

Monte Carlo simulations: A primer

D. Frenkel

*FOM Institute for Atomic and Molecular Physics,
Kruislaan 407, 1098 SJ Amsterdam, The Netherlands*

1. Introduction

It is easy to understand the Molecular Dynamics (MD) technique, even if you have no background in statistical mechanics. In contrast, some knowledge of statistical mechanics is required to understand the Monte Carlo (MC) method. In this chapter, I will introduce the MC method, assuming that the reader has no knowledge whatsoever of statistical mechanics. Moreover, it is not my aim to rederive all of this vast subject in a few pages. Rather, I shall make use of analogies, intuitive arguments and known results, to arrive at the desired results. This approach makes the present chapter a nightmare for the purist: I will cut many corners, I will declare certain unproven results to be intuitively obvious, and I will even make statements that are not always true. My aim, in this chapter, is twofold. First, I hope to achieve that, at the end of the chapter, the reader will have working knowledge of the MC method. Second, I hope that the reader will not consider the MC method as a black box, but rather as a technique that can be understood and extended in a rational way. I urge those readers who wish to know more about the MC method, or about statistical mechanics in general, to consult the relevant literature. In this context, a series of books on the MC method, edited by K. Binder [1–3], is an important source of information. Although most aspects of MC simulations are discussed in these books, the emphasis is on simulation in the context of statistical physics (e.g. critical phenomena). The book by Allen and Tildesley [4] discusses MC techniques as applied to atomic and molecular systems. A good ‘technical’ description of the MC method as a numerical scheme can be found in a book by Kalos and Whitlock [5]. Finally, quite a few of the articles referred to in the present chapter can be found in an annotated reprint collection, edited by Ciccotti et al. [6]. Those readers who are interested in using the MC method, but not in its ‘derivation’, can skip directly to Sec. 3.

1.1. From MD to MC

The idea behind MD simulations is that we can study the average behavior of a many-particle system simply by computing the natural time evolution of that system numerically and averaging the quantity of interest over a sufficiently long time. To take a specific example, let us consider a fluid consisting of atoms. Suppose that we

wish to compute the average density of the fluid at a distance r from a given atom i , $\rho_i(r)$. Clearly, the instantaneous density depends on the coordinates \mathbf{q}_j of all particles j in the system. As time progresses, the atomic coordinates will change (according to Newton's equations of motion), and hence the density around atom i will change. Provided that we have specified the initial coordinates and momenta of all atoms $\mathbf{q}^N(0), \mathbf{p}^N(0)$ we know, at least in principle, the time evolution of $\rho_i(r; \mathbf{q}^N(0), \mathbf{p}^N(0), t)$. In an MD simulation, we measure the time-averaged density $\overline{\rho_i(r)}$ of a system of N atoms, in a volume V , at a constant total energy E

$$\overline{\rho_i(r)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt' \rho_i(r; t') \quad (1)$$

Note that, in writing down this equation, we have implicitly assumed that for sufficiently long T , the time-average does not depend on the initial conditions. This is, in fact, a subtle assumption that is not true in general (see e.g. Ref. 7). However, as stated above, we shall disregard subtleties and simply assume that, once we have specified N , V and E , time-averages do not depend on the initial coordinates and momenta. If that is so, then we would not change our result for $\overline{\rho_i(r)}$ if we average over many different initial conditions, i.e. we consider the hypothetical situation where we run a large number of MD simulations at the same values for N , V and E , but with different initial coordinates and momenta

$$\overline{\rho_i(r)} = \frac{\sum_{\text{initial conditions}} \left(\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt' \rho_i(r; \mathbf{q}^N(0), \mathbf{p}^N(0), t') \right)}{\text{number of initial conditions}} \quad (2)$$

We now consider the limiting case where we average over *all* initial conditions compatible with the imposed values of N , V and E . In that case, we can replace the sum over initial conditions by an integral

$$\frac{\sum_{\text{initial conditions}} f(\mathbf{q}^N(0), \mathbf{p}^N(0))}{\text{number of initial conditions}} \rightarrow \frac{\int_E d\mathbf{q}^N d\mathbf{p}^N f(\mathbf{q}^N(0), \mathbf{p}^N(0))}{\Omega(N, V, E)} \quad (3)$$

where $\Omega(N, V, E) = \int_E d\mathbf{q}^N d\mathbf{p}^N$ (we have ignored a constant factor, see note [8]). The subscript E on the integrals indicates that the integration is restricted to a shell of constant energy E . We call such an average over 'phase space' (i.e. over all coordinates and momenta), an *ensemble average*, and we denote it by $\langle \dots \rangle$, to distinguish it from a time average, denoted by a bar. The reader may expect that, by now, we are seriously over-averaging and, indeed, this is the case. If we switch the order of the time averaging and the averaging over initial conditions, we find

$$\overline{\rho_i(r)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt' \langle \rho_i(r; \mathbf{q}^N(0), \mathbf{p}^N(0), t') \rangle_{NVE} \quad (4)$$

However, the ensemble average in the above equation does not depend on the time t' . This is so, because there is a one-to-one correspondence between the initial phase-space coordinates of a system, and those that specify the state of the system at a later time t' (see e.g. [7,9]). Hence, averaging over all initial phase-space coordinates is equivalent to averaging over the time-evolved phase space coordinates. For this reason we can leave out the time averaging in Eq. 4 and we find

$$\overline{\rho_i(\mathbf{r})} = \langle \rho_i(\mathbf{r}) \rangle_{NVE} \quad (5)$$

The above equation states that, if we wish to compute the average function of the coordinates and momenta of a many-particle system, we can *either* compute that quantity by time averaging (the 'MD' approach), *or* by ensemble averaging (the 'MC' approach). I should stress that I have only tried to make Eq. 5 *plausible*. I have not proven it. In fact, that would have been quite impossible because Eq. 5 is not true in general. However, in what follows, we shall simply assume that the 'ergodic hypothesis', as Eq. 5 is usually referred to, applies to the systems that we study in computer simulations. The reader should, however, be aware that there are many examples of systems that are not ergodic *in practice* (e.g. glasses, metastable phases) or even *in principle*, e.g. nearly harmonic solids. I refer the reader who is interested in the more fundamental aspects of statistical mechanics to the excellent textbooks on this subject, for instance Refs. 7 and 10–12.

1.2. From ensembles to sampling

Although we have now expressed the average value of a measurable quantity in terms of an integral over phase space, Eq. 5 is never used in practice to compute such averages. In fact, two steps are needed to transform Eq. 5 into a form that is convenient for numerical simulation. The first step, although useful, is not essential, but the second is. Let us first consider the non-essential change. It is often (though not always, see Ref. 13) convenient to perform MC simulations on a system of constant *temperature* T , rather than at constant total *energy* E . Below, I shall show how averages at constant N , V and T can be expressed. The second, non-trivial, step is to construct an algorithm that allows us to estimate ensemble averages in a reasonable amount of computing time. It is only at this stage that the MC scheme will be introduced.

1.2.1. From N, V, E to N, V, T

Consider a system of N particles in a volume V . The energy E of that system is a function of the coordinates \mathbf{q}^N and momenta \mathbf{p}^N of all particles in the system. Quite generally, the total energy can be written as a sum of the kinetic energy K and the potential energy U . The latter is a function of the coordinates \mathbf{q}^N only. It is conventional to refer to this sum of the kinetic and potential energy as the 'Hamiltonian' of

D. Frenkel

the system, denoted by the symbol $H(\mathbf{p}^N, \mathbf{q}^N) \equiv K + U$. Clearly, we can express the condition that the energy of the system has a fixed value E , simply as

$$H(\mathbf{p}^N, \mathbf{q}^N) = E$$

The 'volume' in phase space $\Omega(N, V, E)$, that satisfies the condition $H = E$, is given by [14]:

$$\Omega(N, V, E) = \int d\mathbf{p}^N d\mathbf{q}^N \delta(H(\mathbf{p}^N, \mathbf{q}^N) - E) \quad (6)$$

Next we consider what happens if we bring this system in contact with a large heat bath. Let us denote the number of particles in the heat bath by M ($\gg N$), and its volume by V_0 ($> V$). The heat bath can exchange energy with the N -particle system, but we assume that the total energy of the two systems has a fixed value, say E_0 . Let us now compute the average of a property $X(\mathbf{p}^N, \mathbf{q}^N)$ of the small system

$$\langle X \rangle_{NVT} = \frac{\int dE \langle X \rangle \Omega(N, V, E) \Omega(M, V_0, E_0 - E)}{\int dE \Omega(N, V, E) \Omega(M, V_0, E_0 - E)} \quad (7)$$

As $E_0 > E$, we can write

$$\ln \Omega(M, V_0, E_0 - E) = \ln \Omega(M, V_0, E_0) - E \left(\frac{\partial \ln \Omega(M, V_0, E')}{\partial E'} \right)_{E'=E_0} + O(E^2) \quad (8)$$

The terms of order E^2 vanish in the limit $M \rightarrow \infty$. We denote the derivative

$$\frac{\partial \ln \Omega(M, V_0, E')}{\partial E'}$$

by β . Note that, whereas $\ln \Omega$ is an extensive quantity [15], β is intensive. In fact $\ln \Omega$ is directly related to the entropy S of the system, and β to the temperature. I can make this plausible, using the following argument. First of all, we note that Ω is a function that depends on the state of the system, and not on its history: i.e. it is a thermodynamic state function. Secondly, it follows from thermodynamics that if we transfer an amount of energy E from the heat bath to the small system, the entropy of the heat bath changes by an amount $\Delta S = -E/T$. If we insert this relation into Eq. 8, we find that

$$\Delta \ln \Omega = \Delta S \times T \times \beta \quad (9)$$

where we have ignored the terms $O(E^2)$ that vanish in the limit of an infinite heat bath. As both $\ln \Omega$ and S are extensive thermodynamic state functions, it seems plausible to assume that the two are, in fact, linearly related. This implies that $\beta \times T$ equals

a constant, denoted (for historical reasons) by k_B^{-1} (k_B is the Boltzmann constant). Our relation between S and $\ln \Omega$ then becomes

$$k_B^{-1} S = \ln \Omega + \text{constant} \quad (10)$$

In classical statistical mechanics, the absolute value of the entropy is left undetermined. See, however, [8]. Note that definition of the entropy given in Eq. 10 is consistent with our earlier definition of β

$$\beta = \left(\frac{\partial \ln \Omega}{\partial E} \right) = k_B^{-1} \left(\frac{\partial S}{\partial E} \right) = k_B^{-1} \frac{1}{T} \quad (11)$$

where the last quantity is simply the thermodynamic relation between temperature and entropy. We can now use the above identification of the various terms in Eq. 8 to write

$$\Omega(M, V_0, E_0 - E) = \Omega(M, V_0, E_0) \times \exp[-\beta E]$$

If we insert this result in Eq. 7, we find

$$\langle X \rangle_{NVT} = \frac{\int dE \langle X \rangle \Omega(N, V, E) \exp[-\beta E]}{\int dE \Omega(N, V, E) \exp[-\beta E]} \quad (12)$$

Using Eq. 6, and the fact that

$$\int dE f(E) \delta(H(\mathbf{p}^N, \mathbf{q}^N) - E) = f(H(\mathbf{p}^N, \mathbf{q}^N))$$

we can rewrite Eq. 12 as

$$\langle X \rangle_{NVT} = \frac{\int d\mathbf{q}^N d\mathbf{p}^N X(\mathbf{q}^N, \mathbf{p}^N) \exp[-\beta H(\mathbf{p}^N, \mathbf{q}^N)]}{\int d\mathbf{q}^N d\mathbf{p}^N \exp[-\beta H(\mathbf{p}^N, \mathbf{q}^N)]} \quad (13)$$

The denominator in Eq. 13 is (but for a constant factor, see Eq. 36 below) equal to the so-called partition function, Q , of the system. If we know the partition function of the system, we can compute the Helmholtz free energy. Such calculations constitute a separate topic that will be discussed in detail in later chapters. Having written down Eq. 13, the problem that now remains is to actually *compute* the desired average $\langle X \rangle_{NVT}$. This is where the MC scheme enters the scene. Before we proceed, we make one more simplification. In Eq. 13 the observable X has been expressed as a function of coordinates and momenta. As the kinetic energy is a quadratic function of the momenta, the integration over momenta can be carried out analytically. Hence,

averages of functions that depend on momenta only are usually easy to evaluate. The difficult problem is the computation of averages of functions $X(q^N)$. Only in a few exceptional cases can the multidimensional integral over particle coordinates be computed analytically; in all other cases numerical techniques must be used.

1.3. Why MC sampling?

Monte Carlo sampling is a numerical technique that can be used to evaluate multidimensional integrals of the type discussed in the previous section. It might appear that the most straightforward approach would be to evaluate $\langle X \rangle$ in Eq. 13 by numerical quadrature, for instance using Simpson's rule. It is easy to see, however, that such a method is completely useless even if the number of independent coordinates dN (d is the dimensionality of the system) is still very small ($O(10)$). Suppose that we plan to carry out the quadrature by evaluating the integrand on a mesh of points in the dN -dimensional configuration space. Let us assume that we take m equidistant points along each coordinate axis. The total number of points at which the integrand must be evaluated is then equal to m^{dN} . For all but the smallest systems this number becomes astronomically large, even for small values of m . For instance, if we take 100 particles in three dimensions, and $m = 5$, then we would have to evaluate the integrand at 10^{210} points! Computations of such magnitude cannot be performed in the known universe. And this is fortunate, because the answer that would be obtained would have been subject to large statistical errors. After all, numerical quadratures work best on functions that are 'smooth' over distances corresponding to the mesh size. But for most potential energy functions $U(q^N)$, the Boltzmann factor in Eq. 13 is a rapidly varying function of the particle coordinates. Hence an accurate quadrature requires a small mesh spacing (i.e. a large value of m). Moreover, when evaluating the integrand for a dense fluid (say), we would find that for the overwhelming majority of points this Boltzmann factor is vanishingly small. For instance, for a fluid of 100 hard spheres at the freezing point, the Boltzmann factor would be non-zero for 1 out of every 10^{260} configurations!

The above example clearly demonstrates that better numerical techniques are needed to compute thermal averages. One such technique is the MC method, or more precisely, the MC importance-sampling scheme introduced in 1954 by Metropolis et al. [16]. The application of this method to the numerical simulation of dense molecular systems is the subject of the present chapter.

2. The Metropolis method

The closing lines of the previous section suggest that it is in general not possible to evaluate an integral $\int dq^N \exp[-\beta U(q^N)]$ by direct quadrature. However, unless we wish to compute the free energy of the system, we are not interested in the individual values of the integrals in the numerator and denominator of Eq. 13, but rather in the *ratio* of these integrals

$$\langle A \rangle_{\text{NVT}} = \frac{\int dq^N A(q^N) \exp[-\beta U(q^N)]}{\int dq^N \exp[-\beta U(q^N)]} \quad (14)$$

What Metropolis et al. [16] showed is that it is possible to devise an efficient scheme to sample such a ratio [17]. In order to understand the Metropolis method, let us first look more closely at the structure of Eq. 14. In what follows we denote the configurational part of the partition function by Z

$$Z \equiv \int dq^N \exp[-\beta U(q^N)] \quad (15)$$

Note that the ratio $\exp[-\beta U]/Z$ in Eq. 14 is the probability density to find the system in a configuration around q^N .

Let us denote the probability density by $P(q^N)$. Clearly, $P(q^N)$ is non-negative. Suppose now that we are somehow able to randomly generate points in configuration space according to this probability function $P(q^N)$. This means that, on the average, the number of points n_i generated per unit volume around a point q_i^N is equal to $LP(q_i^N)$, where L is the total number of points that we have generated. In other words

$$\langle A \rangle \approx \frac{1}{L} \sum_{i=1}^L n_i A(q_i^N) \quad (16)$$

Let us next consider how to generate points in configuration space with a relative probability proportional to the Boltzmann factor. The general approach is first to prepare the system under study in a configuration q_i^N that has a non-vanishing Boltzmann factor $\rho(q_i^N)$. This configuration may, for example, correspond to a regular crystalline lattice with no hard-core overlaps. Next we generate a new trial configuration q_j^N by adding a small random displacement δq^N to q_i^N . The Boltzmann factor of this trial configuration is $\rho(q_j^N)$. We must now decide whether we will accept or reject the trial configuration. There are many possible rules for making this decision that satisfy the constraint that on the average the probability of finding the system in a configuration q_i^N is proportional to $w(q_i^N)$. Here we discuss only the Metropolis scheme, because it is simple and generally applicable.

Let us now 'derive' the Metropolis scheme to determine the transition probability π_{ij} to go from configuration q_i^N to q_j^N . It is convenient to start with a thought experiment (actually a thought simulation). We carry out a very large number (say M) of MC simulations in parallel, where M is much larger than the total number of accessible configurations. We denote the number of points in any configuration q_i^N by m_i . We wish that, on average, m_i is proportional to $\rho(q_i^N)$. There is one obvious condition that the matrix elements π_{ij} must satisfy: namely that they do not destroy such an equilibrium distribution once it is reached. This means that, in equilibrium, the average number of accepted trial moves that result in the system leaving state i must be exactly the average number of accepted trial moves from all other states j to state i . It is convenient to impose a much stronger condition, namely that in equilib-

D. Frenkel

rium the average number of accepted moves from i to any one state j is exactly cancelled by the number of reverse moves. This 'detailed balance' condition implies the following

$$\rho_i \pi_{ij} = \rho_j \pi_{ji} \quad (17)$$

There are many possible forms of the transition matrix π_{ij} that satisfy Eq. 17. Let us look how π_{ij} is constructed in practice. We recall that an MC 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} . We now *choose* α to be a symmetric matrix. With this choice for α , we can rewrite Eq. 17 in terms of P_{ij}

$$\rho_i P_{ij} = \rho_j P_{ji} \quad (18)$$

From Eq. 18 follows

$$\frac{P_{ij}}{P_{ji}} = \frac{\rho_j}{\rho_i} = \exp[-\beta(U_j - U_i)] \quad (19)$$

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. is

$$P_{ij} = \begin{cases} \rho_j/\rho_i & \rho_j < \rho_i \\ 1 & \rho_j \geq \rho_i \end{cases} \quad (20)$$

Other choices for P_{ij} are possible (for a discussion, see for instance [4]), but the original choice of Metropolis et al. appears to result in a more efficient sampling of configuration space than most of the other strategies that have been proposed.

In summary, then, in the Metropolis scheme, the transition probability for going from state i to state j is given by

$$\pi_{ij} = \begin{cases} \alpha_{ij} & \rho_j \geq \rho_i \\ \alpha_{ij}(\rho_j/\rho_i) & \rho_j < \rho_i \end{cases} \quad (21)$$

$$\pi_{ii} = 1 - \sum_{j \neq i} \pi_{ij}$$

Note that we still have not specified the matrix α , except 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 the next section. One thing that we have

not yet explained is how to decide whether a trial move is to be accepted or rejected. The usual procedure is as follows: Suppose that we have generated a trial move from state i to state j , with $U_j > U_i$. According to Eq. 19 this trial move should be accepted with a probability $P_{ij} = \exp[-\beta(U_j - U_i)] < 1$. In order to decide whether to accept or reject the trial move, we generate a random number r from a uniform distribution in the interval $[0,1]$. Clearly, the probability that r is less than P_{ij} is equal to P_{ij} . We now accept the trial move if $r < P_{ij}$ and reject it otherwise. This rule guarantees that the probability to accept a trial move from i to j is indeed equal to P_{ij} . Obviously it is very important that our random number generator does indeed generate numbers uniformly in the interval $[0,1]$. Otherwise the MC sampling will be biased. The quality of random number generators should never be taken for granted. Good discussions of random number generators can be found in the book 'Numerical Recipes' [18] and in the book on MC techniques by Kalos and Whitlock [5].

To conclude this rather formal introduction to the Metropolis MC method, let me conclude with a simple example that illustrates the difference between quadrature and importance sampling. Suppose that a group of concerned citizens wishes to determine what fraction of the national high-voltage grid runs in the vicinity of houses. Unfortunately, they do not know exactly how many high-voltage lines there are, nor where they are located. They only know that all high-voltage lines are connected. The conventional ('quadrature') method to measure the density of houses near a power line (say, within 75 m), would be to set up a grid of equidistant mesh points covering the entire country. For every grid point, one would first determine if it is within 75 m of a power line and secondly if the mesh point is located within a house. Clearly, as most of the countryside is *not* near a power line, this scheme is very inefficient. The 'Metropolis' approach would be to start a random walk anywhere near a power line. In this random walk, a trial move is rejected if it takes you more than 75 m from the nearest power line, and is accepted otherwise. After *every* trial move (accepted or not), we add one to a counter if we are inside a house, and zero otherwise. The (unweighted) average of all these measurements yields an estimate for the fraction of land near a power line that is covered by houses. This is the essence of the Metropolis method.

3. Trial moves

Now that we have specified the general structure of the Metropolis algorithm, we should consider its implementation. As the problem of selecting intermolecular potentials for the model system under study is discussed in Sec. 2, I shall simply assume that we have an atomic or molecular system in a suitable starting configuration, and that we have specified all intermolecular interactions. We must now set up the underlying Markov chain, i.e. the matrix α . In more down-to-earth terms: we must decide how we are going to generate trial moves. We should distinguish between trial moves that involve only the molecular centers of mass and those that change the orientation or possibly even the conformation of a molecule.

3.1. Translational moves

We start our discussion with trial moves of the molecular centers of mass. A perfectly acceptable method to create a trial displacement is to add random numbers between $-\Delta$ and $+\Delta$ to the x , y and z coordinates of the molecular centre of mass

$$\begin{aligned}x_i &\rightarrow x_i + \Delta (r_1 - 0.5) \\y_i &\rightarrow y_i + \Delta (r_2 - 0.5) \\z_i &\rightarrow z_i + \Delta (r_3 - 0.5)\end{aligned}\tag{22}$$

where r_i are random numbers uniformly distributed between 0 and 1. Clearly, the reverse trial move is equally probable (hence α is symmetric). However, we are faced with two questions, namely: how large should we choose Δ ? And: should we attempt to move all particles simultaneously, or rather move one at a time? In the latter case we should pick the molecule that is to be moved at random, to ensure that the underlying Markov chain remains symmetric. All other things being equal, we should choose the most efficient sampling procedure. We use the following definition of an optimal sampling scheme: *An MC sampling scheme can be considered optimal, if it yields the lowest statistical error in the quantity to be computed for a given expenditure of 'computing budget'*. Usually, 'computing budget' is equivalent to CPU time.

From this 'definition' it is clear that, in principle, a sampling scheme may be optimal for one quantity but not for another. Actually, the above definition is all but useless in practice (as are most definitions). It is just not worth the effort to measure the error estimate in, for instance, the pressure, for a number of different MC sampling schemes in a series of runs of fixed length. However, it is reasonable to assume that the mean-square error in the 'observable' is inversely proportional to the number of 'uncorrelated' configurations visited in a given amount of CPU time. And the number of independent configurations visited is a measure for the distance covered in configuration space. This suggests a more manageable criterion to estimate the efficiency of an MC sampling scheme: namely the sum of the squares of all accepted trial displacements divided by the computing time. This quantity should be distinguished from the mean-square displacement per unit of computing time, because the latter quantity goes to zero in the absence of diffusion (e.g. in a solid or a glass), whereas the former does not.

Using this criterion it is easy to show that for simulations of condensed phases it is usually advisable to move particles one at a time. To see this, let us consider a system of N spherical particles, interacting through a potential energy function $U(\mathbf{r}^N)$. Typically, we expect that a trial move will be rejected if the potential energy of the system changes by much more than $k_B T$. At the same time we try to make the MC trial steps as large as is possible without having a very low acceptance. A displacement that would, on average, give rise to an increase of the potential energy by

$k_B T$ would still have reasonable acceptance. In the case of a single particle trial move we then have

$$\begin{aligned} \langle \Delta U \rangle &= \left\langle \frac{\partial U}{\partial r_i^\alpha} \right\rangle \overline{\Delta r_i^\alpha} + \frac{1}{2} \left\langle \frac{\partial^2 U}{\partial r_i^\alpha \partial r_i^\beta} \right\rangle \overline{\Delta r_i^\alpha \Delta r_i^\beta} + \dots \\ &= 0 + f(u) \overline{\Delta r_i^2} + O(\Delta^4) \end{aligned} \quad (23)$$

where the angular brackets denote averaging over the ensemble, and the horizontal bar denotes averaging over random trial moves. The second derivative of U has been absorbed into the function $f(U)$, the precise form of which does not concern us here. If we now equate ΔU on the right hand side of Eq. 23 to $k_B T$, we find the following expression for $\overline{\Delta r_i^2}$

$$\overline{\Delta r_i^2} \approx k_B T / f(u) \quad (24)$$

If we attempt to move N particles, one at a time, most of the computation involved is spent on the evaluation of the change in potential energy. Assuming that we use a neighbor list or a similar time saving device, the total time spent on evaluating the potential energy change is proportional to nN , where n is the average number of interaction partners per molecule. The sum of the mean-square displacement will be proportional to $N \overline{\Delta r^2} \sim N k_B T / f(U)$. And hence the mean-square displacement per unit of CPU time will be proportional to: $k_B T / n f(U)$. Now suppose that we try to move all particles at once. The cost of CPU time will still be proportional to nN . But, using the same reasoning as in Eqs. 23 and 24, we estimate that the sum of the mean-square displacements is smaller by a factor $1/N$. Hence the total efficiency will be down by the same factor. This simple argument explains why most simulators use single particle, rather than collective trial moves [19].

Next, consider the choice of the parameter Δ which determines the size of the trial move. How large should Δ be? If it is very large, it is likely that the resulting configuration will have high energy and the trial move will probably be rejected. If it is very small, the change in potential energy is probably small and most moves will be accepted. In the literature one often finds the mysterious statement that an acceptance of approximately 50% should be optimal. This statement is not necessarily true. The optimum acceptance ratio is the one that leads to the most efficient sampling of configuration space. If we express efficiency as mean-square displacements per CPU time, it is easy to see that different MC codes will have different optimal acceptance ratios. The reason is that it makes a crucial difference if the amount of computing required to test whether a trial move is accepted depends on the magnitude of the move (see Fig. 1).

For continuous potentials where all interactions have to be computed before a move can be accepted or rejected, the amount of computation does not depend on the size of the trial move. But the situation is very different for simulations of molecules with hard repulsive cores. Here a move can be rejected as soon as overlap with any neigh-

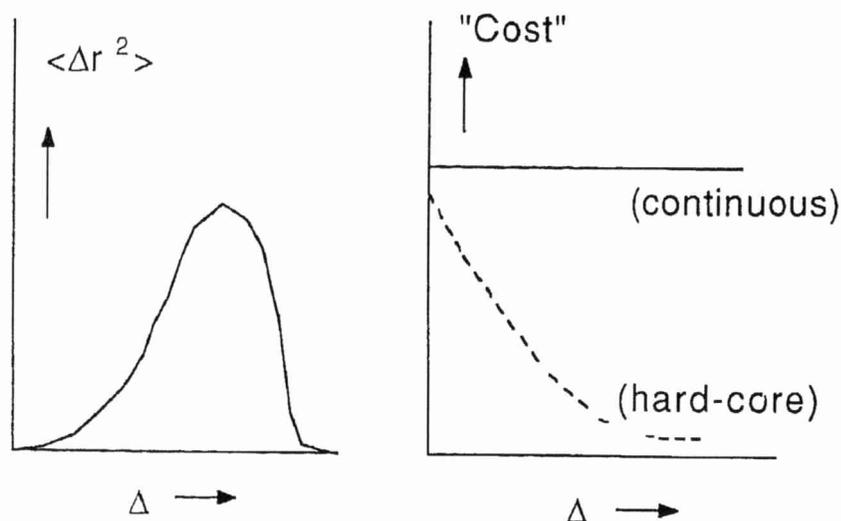


Fig. 1. Left: Typical dependence of the mean-square displacement of a particle on the average size Δ of the trial move. Right: Typical dependence of the computational cost of a trial move on the step size Δ . For continuous potentials, the cost is constant, while for hard-core potentials it decreases rapidly with the size of the trial move.

bor is detected. In this case, a rejected move is cheaper than an accepted one, and hence the average computing time per trial move goes down as the step size is increased. As a result, the optimal acceptance ratio for hard-core systems is appreciably lower than for systems with continuous interactions. Exactly how much depends on the nature of the program, in particular on whether it is a scalar or a vector code (in the latter case hard-core systems are treated much like continuous systems), how the information about neighbor lists is stored, and even on the computational 'cost' of random numbers and exponentiation. The consensus seems to be that for hard-core systems the optimum acceptance ratio is closer to 20% than to 50%, but this is just another rule-of-thumb.

3.2. Orientational moves

If we are simulating molecules rather than atoms we must also generate trial moves that change the molecular orientation. As we suggested in Sec. 3.1, it almost requires an effort to generate translational trial moves with a distribution that does not satisfy the symmetry requirement of the underlying Markov chain. For rotational moves, the situation is very different. It is only too easy to introduce a systematic bias in the orientational distribution function of the molecules by using a non-symmetrical orientational sampling scheme. Several different strategies to generate rotational displacements are discussed in Ref. 4. Here I will only mention one possible approach.

3.2.1. Rigid, linear molecules

Consider a system consisting of N linear molecules. We specify the orientation of the i th molecule by a unit vector \mathbf{u}_i . One possible procedure to change \mathbf{u}_i by a small, random amount is the following. First, we generate a unit vector \mathbf{v} with random orientation. This is quite easy to achieve (and is left as an exercise to the reader; if you wish to check your result, see Ref. 4). Next we multiply this random unit vector \mathbf{v} by a scale factor γ . The magnitude of γ determines the magnitude of the trial rotation. Next we add $\gamma \mathbf{v}$ to \mathbf{u}_i . Let us denote the resulting sum vector by \mathbf{t} : $\mathbf{t} = \gamma \mathbf{v} + \mathbf{u}_i$. Note that \mathbf{t} is not a unit vector. Finally we normalize \mathbf{t} , and the result is our trial orientation vector \mathbf{u}' . We still have to fix γ , which determines the acceptance probability for the orientational move. The optimum value of γ is determined by essentially the same criteria as for translational moves. We have not yet indicated whether or not the translational and orientational moves should be performed simultaneously. Both procedures are acceptable. However, if rotation and translation correspond to separate moves, then the selection of the type of move should be probabilistic rather than deterministic.

3.2.2. Rigid, non-linear molecules

Only slightly more complex is the case of a non-linear rigid molecule. It is conventional to describe the orientation of non-linear molecules in terms of the Eulerian angles (φ, θ, ψ) . However, for most simulations, use of these angles is less convenient because all rotation operations should then be expressed in terms of trigonometric functions, and these are computationally 'expensive'. It is usually better to express the orientation of such a molecule in terms of quaternion parameters (for a discussion of quaternions in the context of computer simulation, see Ref. 4). The rotation of a rigid body can be specified by a quaternion of unit norm \mathbf{Q} . Such a quaternion may be thought of as a unit vector in four-dimensional space

$$\mathbf{Q} \equiv (q_0, q_1, q_2, q_3) \quad \text{with} \quad q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1 \quad (25)$$

There is a one-to-one correspondence between the quaternion components q_α and the Eulerian angles

$$\begin{aligned} q_0 &= \cos \frac{\theta}{2} \cos \left(\frac{\varphi + \psi}{2} \right) \\ q_1 &= \sin \frac{\theta}{2} \cos \left(\frac{\varphi - \psi}{2} \right) \\ q_2 &= \sin \frac{\theta}{2} \sin \left(\frac{\varphi - \psi}{2} \right) \\ q_3 &= \cos \frac{\theta}{2} \sin \left(\frac{\varphi + \psi}{2} \right) \end{aligned} \quad (26)$$

D. Frenkel

and the rotation matrix R , which describes the rotation of the molecule-fixed vector in the laboratory frame, is given by (see e.g. [20])

$$R = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & (q_0^2 - q_1^2 + q_2^2 - q_3^2) & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_2 - q_0 q_3) & 2(q_2 q_3 + q_0 q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix} \quad (27)$$

In order to generate trial rotations of non-linear rigid bodies, we must rotate the vector (q_0, q_1, q_2, q_3) on the four-dimensional unit sphere. The procedure described above for the rotation of a 3D unit vector is easily generalized to 4D. An efficient method to generate random vectors uniformly on the 4D unit sphere has been suggested by Vesely [20].

3.2.3. Non-rigid molecules

The MC schemes that are needed to perform an efficient sampling of the conformations of flexible molecules, are rather different from the schemes discussed above. In fact, as the chapter by Siepmann in this volume is devoted exclusively to the description of such sampling schemes, I shall not discuss them here.

3.3. 'Unphysical' MC moves

An important advantage of the MC scheme is that it is possible to carry out 'unphysical' trial moves that greatly enhance the speed with which configuration space is sampled. An example is provided by the simulation of mixtures. In a *Molecular Dynamics* simulation of, for instance, a binary (A-B) mixture, the efficiency with which configuration space is sampled is greatly reduced by the fact that concentration fluctuations decay very slowly (typically the relaxation time $\tau \sim D_{AB}/\lambda^2$ where D_{AB} is the mutual diffusion coefficient and λ is the wavelength of the concentration fluctuation). This implies that very long runs are needed to ensure equilibration of the local composition of the mixture. In solids, equilibration may not take place at all. In contrast, in a *Monte Carlo* simulation, it is permissible to carry out trial moves that swap the identities of the two particles of the species A and B. Such moves, even if they have only a moderate rate of acceptance (a few percent will do), greatly speed up the sampling of concentration fluctuations. In recent years, much progress has been made in the construction of new 'unphysical' MC moves that can greatly improve the sampling efficiency. Best known among these is the 'Swendsen-Wang' scheme to perform collective moves on groups of coordinates [21] and subsequent extensions and modifications of that scheme (see Ref. 3). Most of these cluster moves are either limited to, or best suited for, lattice systems. For more details, I refer the reader to the original literature.

3.4. 'Physical' MC moves

In addition to the 'unphysical' sampling schemes discussed above, there have been many attempts to construct MC schemes that utilize information about the 'natural' motion of a many-body system to construct 'collective' many-particle moves. It should be stressed that special techniques are required to construct such collective MC moves, because, as was discussed below Eq. 24, 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 MC and MD (see e.g. Ref. 22 and the book of Allen and Tildesley [4]). 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 [23] 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 the preceding chapter by van Gunsteren), 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. 23.

4. Error estimates

Let us consider the measurement of a dynamical quantity A in an MC simulation (the present discussion applies, with minor modifications, to MD simulations). During a simulation of m MC trial moves per particle, we obtain the following estimate for the equilibrium estimate of A

$$A_m = \frac{1}{m} \sum_{i=1}^m A(i) \quad (28)$$

where the subscript on A_m refers to averaging over a finite 'time' m . If the ergodic hypothesis is justified then $A_m \rightarrow A$ as $m \rightarrow \infty$, where $\langle A \rangle$ denotes the ensemble average of A . Let us now estimate the variance in A_m , $\langle \Delta A_m \rangle^2$

$$\begin{aligned} \langle \Delta A_m \rangle^2 &= \langle A_m^2 \rangle - \langle A_m \rangle^2 \\ &= \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \langle (A(i) - \langle A \rangle) (A(j) - \langle A \rangle) \rangle \end{aligned} \quad (29)$$

D. Frenkel

Note that $\langle (A(i) - \langle A \rangle) (A(j) - \langle A \rangle) \rangle$ in Eq. 29 is something like the time-correlation function in fluctuations in the variable A (in fact, for MD simulations it is precisely that). Let us denote this correlation function by $C_A(j-i)$. If the duration of the sampling m is much larger than the characteristic decay time n_A^c of C_A , then we may rewrite Eq. 29 as

$$\langle (\Delta A_m)^2 \rangle \approx \frac{1}{m} \sum_{k=-\infty}^{\infty} C_A(k) \approx \frac{2n_A^c}{m} C_A(0) \quad (30)$$

In the last equation we have used the definition of n_A^c as half the sum from $-\infty$ to ∞ of the normalized correlation function $C_A(k)/C_A(0)$. The relative variance in A_m is therefore given by

$$\frac{\langle (\Delta A_m)^2 \rangle}{\langle A \rangle^2} \approx \frac{2n_A^c}{m} \frac{\langle A^2 \rangle - \langle A \rangle^2}{\langle A \rangle^2} \quad (31)$$

Equation 31 clearly shows that the r.m.s. error in A_m is proportional to $(n_A^c/m)^{1/2}$. This result is hardly surprising. It simply states the well-known fact that the variance in a measured quantity is inversely proportional to the number of uncorrelated measurements. In the present case, this number is clearly proportional to m/n_A^c . This result may be trivial, but is nevertheless very important, because it shows directly how the lifetime and amplitude of fluctuations in an observable A affect the statistical accuracy. This is of particular importance in the study of fluctuations associated with hydrodynamical models or pre-transitional fluctuations near a symmetry breaking phase transition. Such modes usually have a characteristic life time that is proportional to the square of their wavelength. In order to minimize the effects of the finite system size on such phase transitions, it is preferable to study systems with a box size L that is large compared with all relevant correlation lengths in the system. However, due to the slow decay of long-wavelength fluctuations, the length of the simulation needed to keep the relative error fixed should be proportional to L^2 . As the CPU time for a run of fixed length is proportional to the number of particles (at best), the CPU time needed to maintain constant accuracy increases quite rapidly with the linear dimensions of the system (e.g. as L^5 in three dimensions).

There is another aspect of Eq. 31 that is not immediately obvious, namely that it makes a difference whether or not the observable A can be written as a sum of uncorrelated single-particle properties. If this is the case, then it is easy to see that the ratio $(\langle A^2 \rangle - \langle A \rangle^2)/\langle A \rangle^2$ is inversely proportional to the number of particles, N . To see this, consider the expressions for $\langle A \rangle$ and $\langle A^2 \rangle - \langle A \rangle^2$ in this case

$$\langle A \rangle = \sum_{i=1}^N \langle a_i \rangle = N \langle a \rangle \quad (32)$$

and

$$\langle A^2 \rangle - \langle A \rangle^2 = \sum_{i=1}^N \sum_{j=1}^N \langle a_i - \langle a \rangle \rangle \langle a_j - \langle a \rangle \rangle \quad (33)$$

If the fluctuations in a_i are uncorrelated, then we find

$$\frac{\langle A^2 \rangle - \langle A \rangle^2}{\langle A \rangle^2} = \frac{1}{N} \frac{\langle a^2 \rangle - \langle a \rangle^2}{\langle a \rangle^2} \quad (34)$$

From Eq. 34 it is clear that the statistical error in a single-particle property is inversely proportional to $N^{1/2}$. Hence, for single-particle properties the accuracy improves as one goes to larger systems (at fixed length of the simulation). In contrast, no such advantage is to be gained when computing truly collective properties. This is one more reason why bigger is not always better. A more detailed discussion of statistical errors in collective and single-particle time correlation functions can be found in Refs. 24 and 25. Systematical techniques to measure statistical errors in simulations are discussed in Refs. 4 and 26.

5. Other ensembles

In the previous sections we have introduced MC simulation as a technique to compute the average value of an arbitrary function $A(q^N)$ of the coordinates of N particles at constant temperature T and volume V

$$\langle A \rangle_{NVT} = \frac{\int dq^N A(q^N) \exp[-\beta U(q^N)]}{\int dq^N \exp[-\beta U(q^N)]} \quad (35)$$

This average is usually referred to as the ‘canonical’ ensemble average. To be more precise, the true canonical ensemble average (Eq. 13) 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.

Most MC 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.

5.1. Isobaric-isothermal ensembles

The isobaric-isothermal (constant-NPT) ensemble is widely used in MC 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 the 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 MC simulations were first described by Wood [27] 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 [28] 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 MC 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 MC 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

$$Q(N, V, T) \equiv \frac{1}{\Lambda^{3N} N!} \int_0^L \dots \int_0^L d\mathbf{q}^N \exp[-\beta U(\mathbf{q}^N)] \quad (36)$$

where $\Lambda = [h^2/(2\pi m k_B T)]^{1/2}$ is the thermal de Broglie wavelength. It is convenient to rewrite Eq. 36 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 Eq. 36 we obtain

$$Q(N, V, T) \equiv \frac{V^N}{\Lambda^{3N} N!} \int_0^1 \dots \int_0^1 d\mathbf{s}^N \exp[-\beta U(\mathbf{s}^N; L)] \quad (37)$$

In Eq. 37, we have written $U(\mathbf{s}^N; L)$ to indicate that U depends on the *real* rather than the *scaled* distances between the particles. The expression for the Helmholtz free energy of the system is

$$A(N, V, T) = -k_B T \ln Q$$

$$\begin{aligned}
 &= -k_B T \ln \left\{ \frac{V^N}{\Lambda^{3N} N!} \right\} - k_B T \ln \left\{ \int ds^N \exp[-\beta U(s^N; L)] \right\} \\
 &= A_{id}(N, V, T) + A_{cx}(N, V, T)
 \end{aligned} \tag{38}$$

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 plus an excess part. Let us now assume that the system is separated by a 'piston' [29] from an ideal gas reservoir (see Fig. 2). 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

$$Q(N, m, V, V_0, T) = \frac{V^N (V_0 - V)^m}{\Lambda^{3N} N! m!} \int ds^m \int ds^N \exp[-\beta U(s^N; L)] \tag{39}$$

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 $A_{tot} = -k_B T \ln Q(N, M, V, V_0, T)$. Now let us assume that the 'piston' between the two subsystems is free to move, so that the volume V of the N -particle substrate 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 $P(V)$ that the N -particle subsystem has a volume V is given by

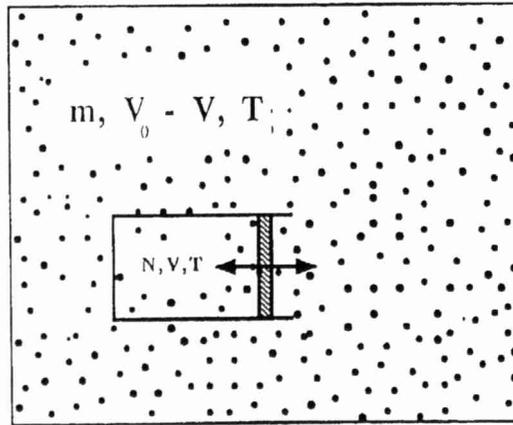


Fig. 2. Ideal gas (m particles, volume $V_0 - V$) can exchange volume with an N -particle system (volume V).

D. Frenkel

$$P(V) = \frac{V^N (V_0 - V)^m \int ds^N \exp[-\beta U(s^N; L)]}{\int_0^{V_0} dV' V'^N (V_0 - V')^m \int ds^N \exp[-\beta U(s^N; L')]} \quad (40)$$

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 of the large system. In other words, the large system works as a manostat for the small system. In that case we can simplify Eqs. 39 and 40. Note that in the limit $V/V_0 \rightarrow 0$ we can write

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

Note that for $m \rightarrow \infty$, $\exp[-mV/V_0] \rightarrow \exp[-\rho V]$. But, as the reservoir contains an ideal gas, ρ can be written as βP . With these substitutions Eq. 40 becomes

$$P(V) = \frac{V^N \exp[-\beta PV] \int ds^N \exp[-\beta U(s^N; L)]}{\int_0^{V_0} dV' V'^N \exp[-\beta PV'] \int ds^N \exp[-\beta U(s^N; L')]} \quad (41)$$

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

$$G(N, P, T) = -k_B T \ln \int dV \frac{V^N \exp[-\beta PV]}{\Lambda^{3N} N!} \int ds^N \exp[-\beta U(s^N; L)] \quad (42)$$

Equation 41 is the starting point for constant-NPT MC 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

$$P(V; s^N) = \frac{V^N \exp[-\beta PV] \exp[-\beta U(s^N; L)]}{\int_0^{V_0} dV' V'^N \exp[-\beta PV'] \int ds^N \exp[-\beta U(s^N; L')]} \quad (43)$$

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

$$\exp[-\beta U(s^N, V) + PV - N\beta^{-1} \ln V]$$

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 symmetry of the underlying Markov chain). Let us assume that our trial moves consist of an attempted change of volume from $V' = V + \Delta V$,

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

$$\exp[-\beta(U(s^N, V') - U(s^N, V) + P(V' - V) - N\beta^{-1} \ln(V'/V))] > R \quad (43)$$

where 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 [28], or in the logarithm of the volume [30]. Such trial moves are equally legitimate, as long as the symmetry of the underlying Markov chain is maintained. However, such alternative schemes result in a slightly different form for Eq. 43.

The frequency with which trial moves in the volume should be attempted is dependent on the efficiency with which 'volume space' is sampled. If, as before, we use as our criterion of efficiency: (*sum of squares of accepted volume changes*)/(*CPU time*), then it is obvious that the frequency with which we attempt moves depends on the 'cost'. In general, a volume trial move will require that we recompute *all* intermolecular interactions. It is therefore comparable in cost to carrying out N trial moves on the molecular positions. In such cases it is common practice to perform one volume trial move for every cycle of positional trial moves. Note that, in order to guarantee the symmetry of the underlying Markov chain, volume moves should not be attempted periodically after a fixed number (say N) positional trial moves. Rather, at every step there should be a probability $1/N$ to attempt a volume move instead of a particle move. The criteria determining the optimal acceptance of volume moves are no different than those for particle moves.

There exists a class of potential energy functions for which volume trial moves are very cheap: namely, those for which the total interaction energy can be written as a sum of powers of the intermolecular distances

$$\begin{aligned} U_n &= \sum_{i < j} \varepsilon \left(\frac{\sigma}{r_{ij}} \right)^n \\ &= \sum_{i < j} \varepsilon \left(\frac{\sigma}{L s_{ij}} \right)^n \end{aligned} \quad (44)$$

or, possibly, a linear combination of such sums (the famous Lennard-Jones potential is an example of the latter category). Note that U_n in Eq. 44 changes in a trivial way if the volume is modified such that the linear dimensions of the system change from L to L'

$$U_n(L') = \left(\frac{L}{L'} \right)^n U_n(L) \quad (45)$$

Clearly, in this case, computing the probability of acceptance of a volume changing trial move is extremely cheap. Hence such trial moves may be attempted with high

D. Frenkel

frequency, e.g. as frequent as particle moves. One should be careful when using the scaling property (Eq. 45) if one uses at the same time a cutoff (say r_c) on the range of the potential. Use of Eq. 45 implicitly assumes that the cutoff radius r_c scales with L , such that $r_c' = r_c(L'/L)$. The corresponding tail correction to the potential (and the virial) should also be recomputed to take into account both the different cutoff radius and the different density of the system. Finally, it is always useful to compute the virial pressure during a constant pressure simulation. On average, the virial pressure should always be equal to the applied pressure. This is easy to prove as follows. First of all, note that the virial pressure $P_v(V)$ of an N -particle system at volume V is equal to

$$P_v(V) = - \left(\frac{\partial A}{\partial V} \right)_{NT} \quad (46)$$

In an isothermal-isobaric ensemble the probability density $P(V)$ of finding the system with volume V is equal to $\exp[-\beta(A(V) + PV)]/Z_{NPT}$, where

$$Z_{NPT} \equiv \int dV \exp[-\beta(A(V) + PV)]$$

Let us now compute the average value of the virial pressure

$$\begin{aligned} \langle P_v \rangle &= \frac{\int dV \left(- \frac{\partial A(V)}{\partial V} \right) \exp[-\beta(A(V) + PV)]}{Z_{NPT}} \\ &= \frac{\int dV \beta^{-1} \left(\frac{\partial \exp[-\beta A(V)]}{\partial V} \right) \exp[-\beta PV]}{Z_{NPT}} \\ &= \frac{\int dV P \exp[-\beta(A(V) + PV)]}{Z_{NPT}} \\ &= P \end{aligned} \quad (47)$$

The third line in this equation follows from the partial integration.

Thus far we have limited our discussion of constant pressure MC to pure atomic systems. Extension of the technique to mixtures is straightforward. The method is also easily applicable to molecular systems. However, in the latter case it is crucial to note that *only the center of mass positions of the molecules should be scaled in a volume move* but never the relative positions of the constituent atoms in the molecule. This has one practical consequence, namely that the simple scaling relation (Eq. 45) can never be used in molecular systems with site-site interactions. The reason is that, even if the center of mass separations between molecules scale as a simple power of the system size, the site-site separations do not.

5.1.1. Isotension-isothermal ensemble

The NPT-MC method is perfectly adequate for homogeneous fluids. However, for inhomogeneous systems, in particular crystalline solids, it may not be sufficient that the simulation box can change size. Often we are interested in the transformation of a crystal from one structure to another, or even in the change of shape of the crystalline unit cell with temperature or with applied stress. In such cases it is essential that there is enough freedom in the shape of the simulation box to allow for such changes in crystal structure without creating grain boundaries or highly stressed configurations. This problem was first tackled by Parrinello and Rahman [31,32] who developed an extension of the constant-pressure MD technique introduced by Andersen [33]. The extension of the Parrinello-Rahman method to MC simulations is straightforward (actually, the method is quite a bit simpler in MC than in MD). To my knowledge, the first published account of constant 'stress' MC is a paper by Najafabadi and Yip [34] to which I refer the reader for more details.

5.2. Grand canonical ensemble

In Sec. 2, we introduced the Metropolis sampling scheme as a method to compute thermal averages of functions $A(q^N)$, 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 could not be used to determine the integral $\int dq^N \exp[-\beta U(q^N)]$ itself. The latter quantity measures the effective volume in configuration space that is accessible to the system. Hence the original Metropolis scheme could not 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 A , 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) developed and first implemented for classical fluids by Norman and Filinov [35], and later extended and improved by a number of other groups [36–44]. The basic idea of the GCMC method is explained below.

In order to understand the statistical mechanical basis for the GCMC technique, let us return to Eq. 39 of Sec. 5.1. 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 $V_0 - V$

$$Q(N, m, V, V_0, T) = \frac{V^N (V_0 - V)^m}{\Lambda^{3N} N! m!} \int ds^m \int ds^N \exp[-\beta U(s^N; L)]$$

Now, instead of allowing the two systems to exchange volume, let us see what happens if the systems can exchange particles (see Fig. 3).

To be more 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 U changes from $U(s^N)$ to $U(s^{N+1})$. The expression for the total partition function of the system, including all possible distributions of the M particles over the two sub-volumes is

$$Q(M, V, V_0, T) = \sum_{N=0}^M \frac{V^N (V_0 - V)^{M-N}}{\Lambda^{3M} N! (M-N)!} \int ds^{M-N} \int ds^N \exp[-\beta U(s^N)] \quad (48)$$

Following the approach of Sec. 5.1, 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

$$P(s^M; N) = \frac{V^N V'^{M-N}}{Q(M, V, V', T) \Lambda^{3M} N! (M-N)!} \exp[-\beta U(s^N)] \quad (49)$$

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 (Eq. 49)

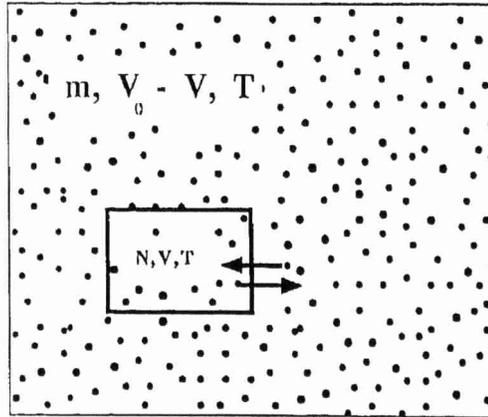


Fig. 3. Ideal gas (m particles, volume $V_0 - V$) can exchange particles with an N -particle system (volume V).

$$P_{N \rightarrow N+1} = \frac{V(M-N)}{V'(N+1)} \exp[-\beta(U(\mathbf{s}^{N+1}) - U(\mathbf{s}^N))] \quad (50)$$

$$P_{N+1 \rightarrow N} = \frac{V'(N+1)}{V(M-N)} \exp[-\beta(U(\mathbf{s}^N) - U(\mathbf{s}^{N+1}))] \quad (51)$$

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 $(M/N) \rightarrow \infty$, Eqs. 50 and 51 become

$$P_{N \rightarrow N+1} = \frac{V}{N+1} \Lambda^{-3} \exp[\beta \mu] \exp[-\beta(U(\mathbf{s}^{N+1}) - U(\mathbf{s}^N))] \quad (52)$$

$$P_{N+1 \rightarrow N} = \frac{N+1}{V} \Lambda^3 \exp[-\beta \mu] \exp[-\beta(U(\mathbf{s}^N) - U(\mathbf{s}^{N+1}))] \quad (53)$$

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

Equations 52 and 53 are the basic equations for Grand Canonical MC 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 MC sampling. Note that in Eqs. 52 and 53 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 $\langle U \rangle$. 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 Sec. 2 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 the 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 [42]. 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. For more details on the Grand Canonical MC method, the reader is referred to the book by Allen and Tildesley [4] and a review paper by Frenkel [45].

5.3. Gibbs ensemble

The conditions 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- μ PT ensemble'. The quotation marks around the name of this 'ensemble' are international, 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 MC the volume V of the system is fixed.

It was, however, shown by Panagiotopoulos [46] that an MC scheme can be devised 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 so-called 'Gibbs-ensemble 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. Below we shall show how the Gibbs method can be derived using the description developed in the previous sections.

We start our discussion by recalling the expression for the partition function for a system of M particles distributed over two volumes V and $V' = V_0 - V$, i.e. Eq. 48 in Sec. 5.2

$$Q(M, V, V', T) \equiv \sum_{N=0}^M \frac{V^N (V')^{M-N}}{\Lambda^{3M} N! (M-N)!} \int ds^{M-N} \int ds^N \exp[-\beta U(s^N)]$$

In the previous section we assumed that the particles in volume V' behaved as ideal gas molecules. Now we consider the case that the particles in both volumes are subject to the same intermolecular interactions. In that case, the partition function becomes

$$Q(M, V, V', T) \equiv \sum_{N=0}^M \frac{V^N (V')^{M-N}}{\Lambda^{3M} N! (M-N)!} \times$$

$$\times \int ds^{M-N} \exp[-\beta U] s^{M-N} \int ds^N \exp[-\beta U(s^N)] \quad (54)$$

We now allow trial moves to transport particles from one sub-system to the other (see Fig. 4). Using the results of the previous section, it is easy to see that the probability ratio associated with a trial move that transfers one particle from volume V' to volume V is given by

$$P_{N \rightarrow N+1} = \frac{V(M-N)}{V'(N+1)} \exp[-\beta((U(s^{N+1}; V) - U(s^N; V) + U(s^{M-N-1}; V') - U(s^{M-N}; V')))] \quad (55)$$

The corresponding expression for the reverse trial move follows by making the appropriate substitutions in Eq. 55. As before, the acceptance or rejection of this trial move can be decided by comparing $P_{N \rightarrow N+1}$ with a random number in the interval $[0,1]$. At this stage, we could set up a MC simulation in which we allow trial moves that transfer particles from V to V' and *vice versa*. In the course of such a simulation, the distribution of particles over V and V' would become such that both sub-systems are at the same chemical potential.

This may correspond to a situation where both sub-systems have the same density and pressure, but this need not be the case. For example, if the overall density of the total system (M/V_0), corresponds to a point in the two-phase region of a first order phase transition, the two sub-systems may end up in different phases (which is de-

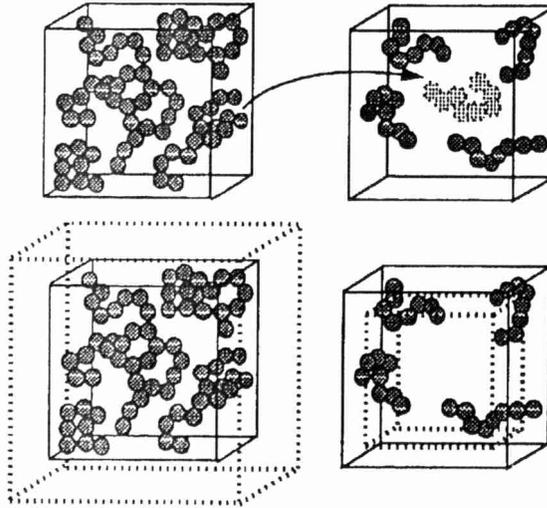


Fig. 4. Gibbs ensemble. Two systems can exchange both volume and particles, but the total volume and the total number of particles is fixed.

sirable), at different pressures (which is undesirable). In order to ensure that the two sub-systems are always at the same pressure, the Gibbs method also allows for volume changing moves. These may either keep the total volume V_0 fixed (this was the formulation of the original paper of Panagiotopoulos [46]), or the two systems may both be in contact with a constant pressure bath at pressure P (this generalization is described in a subsequent paper by Panagiotopoulos et al. [47]). Both methods follow directly from the discussion in Sec. 5.1. For the sake of completeness we write down the expressions for the probability ratios associated with a volume changing move in the constant- V_0 ensemble

$$\begin{aligned}
 P(V \rightarrow V + \Delta V, V' \rightarrow V' - \Delta V) = \\
 \exp[-\beta(U_N(V + \Delta V) - U_N(V) + U_{M-N}(V' - \Delta V) - U_{M-N}(V'))] \times \\
 \exp[N \ln((V + \Delta V)/V) + (M - N) \ln((V' - \Delta V)/V')] \quad (56)
 \end{aligned}$$

In the constant- P ensemble volume changing moves on the two subsystems can be carried out independently. The probability of acceptance of such trial moves is given directly by Eq. 43 of Sec. 5.1. Of course, this constant- P method can only be applied to systems containing two or more components because in a one-component system, the two-phase region is a *line* in the P - T -plane. Hence, the probability that any specific choice of P and T will actually be *at* the phase transition, is vanishingly small. In contrast, for two-component systems, the two-phase region corresponds to a finite *area* in the P - T -plane.

Note that in either formulation of the Gibbs method, the total number of particles is fixed. The method can be extended to study inhomogeneous systems [48] and is particularly suited to study phase equilibria in multicomponent mixtures and equilibria across semi-permeable membranes [47]. The great advantage of the Gibbs method over the conventional technique to study phase coexistence is that in the Gibbs method the system spontaneously 'finds' the densities and compositions of the coexisting phases. In contrast, the conventional approach was, for instance, to compute the relevant chemical potentials as a function of pressure at a number of different compositions (this might require an appreciable number of simulations), and then *construct* the coexistence line. At the beginning of this section I stated that the Gibbs ensemble is *not* a 'constant- μ - P - T ' ensemble, but I did not say what ensemble it actually corresponds to. This point has been considered by Smit et al. [49] who concluded that, in the thermodynamic limit, the (constant- V) 'Gibbs' ensemble is rigorously equivalent to the canonical 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 phases 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 [50]. A description of the latter scheme is, however, outside the scope of the present chapter. For more details, the reader is referred to Ref. 50.

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. Binder, K., Monte Carlo Methods in Statistical Physics, Springer, Berlin, 1979.
2. Binder, K., Applications of the Monte Carlo Method in Statistical Physics, Springer, Berlin, 1984.
3. Binder, K., The Monte Carlo Method in Condensed Matter Physics, Springer, Berlin, 1992.
4. Allen, M.P. and Tildesley, D.J., Computer Simulations of Liquids, Clarendon Press, Oxford, 1987.
5. Kalos, M.H. and Whitlock, P.A., Monte Carlo Methods, John Wiley and Sons, New York, NY, 1987.
6. Ciccotti, C., Frenkel, D. and McDonald, I.R., Simulation of Liquids and Solids, North-Holland, Amsterdam, 1987.
7. Reichl, L.E., A Modern Course in Statistical Physics, University of Texas Press, Austin, TX, 1980.
8. If we consider a quantum-mechanical system, then $\Omega(N, V, E)$ is simply the number of quantum states of that system, for given N , V and E . In the classical limit, the number of quantum states of a d -dimensional system of N distinguishable, structureless particles is given by $\Omega(N, V, E) = \int dp^N dq^N / h^{dN}$. For N indistinguishable particles, we should divide the latter expression by $N!$.
9. Goldstein, H., Classical Mechanics, 2nd ed., Addison Wesley, Reading, MA, 1980.
10. Feynman, R.P., Statistical Physics, Benjamin, Reading, MA, 1972.
11. Landau, L.D. and Lifshitz, E.M., Statistical Physics, 3rd ed., Pergamon Press, London, 1980.
12. Chandler, D., Introduction to Modern Statistical Mechanics, Oxford University Press, New York, NY, 1987.
13. Creutz, M., Phys. Rev. Lett., 50(1983)1411.
14. In fact, the usual expression for $\Omega(N, V, E)$ contains an additional factor (Jacobian) that corrects for the fact that the δ -function is expressed in terms of the energy, rather than in the phase space coordinates. I ignore this factor as it does not appear in the subsequent expressions.
15. Extensive quantities scale with the system size, intensive quantities are independent of system size. If we combine two systems 1 and 2, with phase space volumes Ω_1 and Ω_2 , then the phase space volume of the combined systems equals $\Omega_1 \times \Omega_2$. This implies that $\ln \Omega_{12} = \ln \Omega_1 + \ln \Omega_2$. In other words, Ω is extensive.

16. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.H., Teller, A.H. and Teller, E., J. Chem. Phys., 21(1953)1087.
17. An interesting account of the early history of the Metropolis method may be found in: Andersen, H.L., J. Stat. Phys., 43(1986)2384; and in Wood, W.W., In Ciccotti, G. and Hoover, W.G. (Eds.) Molecular Dynamics Simulations of Statistical Mechanical Systems, North-Holland, Amsterdam, 1986, p. 3.
18. Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T., Numerical Recipes, Cambridge University Press, Cambridge, 1986.
19. The situation is of course different for random trial moves that affect not all particles, but only a non-interacting subset. In that case the acceptance or rejection can be decided for each particle individually. Such a procedure is in fact very well suited for Monte Carlo simulations on both vector and parallel computers.
20. Vesely, F.J., J. Comput. Phys., 47(1982)291.
21. Swendsen, R.H. and Wang, J.S., Phys. Rev. Lett., 58(1987)86.
22. Rosky, P.J., Doll, J.D. and Friedmann, H.L., J. Chem. Phys., 69(1978)4628.
23. Duane, S., Stevenson, A.D., Pendleton, B.J. and Roweth, D., Phys. Lett., B195(1987)216.
24. Zwanzig, R. and Ailawadi, N.K., Phys. Rev., 182(1969)280.
25. Frenkel, D., In van Kranendonk, J. (Ed.) Intermolecular Spectroscopy and Dynamical Properties of Dense Systems, Soc. Italiana di Fisica, Bologna, 1980, pp. 156-201.
26. Jacucci, G. and Rahman, A., Nuovo Cimento, D4(1984)341.
27. Wood, W.W., J. Chem. Phys., 48(1968)415.
28. McDonald, I.R., Mol. Phys., 23(1972)41.
29. Actually, there is no need to assume a *real* Piston. The systems with volume V and $V_0 - V$ may both be isolate 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 .
30. Eppenga, R. and Frenkel, D., Mol. Phys., 52(1984)1303.
31. Parrinello, M. and Rahman, A., Phys. Rev. Lett., 45(1980)1196.
32. Parrinello, M. and Rahman, A., J. Appl. Phys., 52(1981)7182.
33. Andersen, H.C., J. Chem. Phys., 72(1980)2384.
34. Najafabadi, R. and Yip, S., Scripta Metall., 17(1983)1199.
35. Norman, G.E. and Filinov, V.S., High Temp. Res. USSR, 7(1969)216.
36. Adams, D.J., Mol. Phys., 28(1974)1241.
37. Adams, D.J., Mol. Phys., 29(1975)307.
38. Adams, D.J., Mol. Phys., 32(1976)647.
39. Adams, D.J., Mol. Phys., 37(1979)211.
40. Rowley, L.A., Nicholson, D. and Parsonage, N.G., J. Comput. Phys., 17(1975)401.
41. Yao, J., Greenkorn, R.A. and Chao, K.C., Mol. Phys., 46(1982)587.
42. Mezei, M., Mol. Phys., 40(1980)901.
43. Valleau, J.P. and Cohen, K.L., J. Chem. Phys., 72(1980)3935.
44. Van Meegen, W. and Snook, I., J. Chem. Phys., 73(1980)4656.
45. Frenkel, D., In Ciccotti, G. and Hoover, W.G. (Eds.) Molecular Dynamics Simulations of Statistical Mechanical Systems, North-Holland, Amsterdam, 1986, pp. 151-188.
46. Panagiotopoulos, A.Z., Mol. Phys., 61(1987)813.
47. Panagiotopoulos, A.Z., Quirke, N., Stapleton, M. and Tildesley, D.J., Mol. Phys., 63(1988)527.
48. Panagiotopoulos, A.Z., Mol. Phys., 62(1987)701.
49. Smit, B., de Smedt, Ph. and Frenkel, D., Mol. Phys., 68(1989)931.
50. Kofke, D.A. and Glandt, E.D., Mol. Phys., 64(1988)1105.