JOURNAL OF COMPUTATIONAL PHYSICS 62, 441–463 (1986)

The Convergence Properties and Stochastic Characteristics Inherent in Force-Biased and in Metropolis Monte Carlo Simulations on Liquids

SAUL GOLDMAN

The Guelph–Waterloo Centre for Graduate Work in Chemistry, Guelph Campus, Guelph, Ontario N1G 2W1, Canada

Received August 16, 1983; revised November 19, 1984

The very large number of states arise in a typical Monte Carlo simulation of a liquid was replaced by a relatively small number of states by discretizing the energy distribution function. We found that the subdominant eigenvalues of transition matrices defined over these states only weakly depend on the number of states used in the representation. Transition matrices defined over distribution functions with 276 states were then used to approximate the true transition matrices underlying the force-bias and Metropolis Monte Carlo algorithms. We examine the influence of limited state-to-state accessibility, and of distortion in the force-bias biasing function, on a variety of properties, most notably the subdominant eigenvalues of the transition matrices and the stochastic characteristics of the developing distribution function. Our results suggest that because of the inevitable presence of both limited accessibility and distortion in the biasing function, the force-biased algorithm with a moderate degree of biasing is best.

INTRODUCTION

Over the past five years we have seen the development of a number of algorithms [1–6] whose purpose is to improve the convergence rate of Monte Carlo simulations, relative to rates that are obtainable with the now classical Metropolis method [7]. The force-biased method developed by Berne and his colleagues [1–3], is one of the most widely applied of these methods. Unlike the Metropolis method, in which attempted moves with trial particles are made in random directions, in force-biasing, trial moves are biased in the direction of the force vector and (if nonzero) the torque-vector on the particle. This bias is accounted for in the acceptance–rejection step of the algorithm through the condition for detailed balance.

Detailed balance ensures convergence to the correct distribution function in the limit of an infinite number of steps. An actual simulation is of course terminated after some finite number of steps, with the assumption that the results obtained are sufficiently close to their limiting values for the purpose at hand. Because of finite termination, there will in principle always be a discrepancy (which may or may not

0021-9991/86 $3.00

be numerically significant) between the results obtained and their true values. Since different algorithms sample configuration space differently, we can, at least in principle, expect systematic differences upon finite termination depending on the algorithm used. In a previous article [6], it was found that the force-biased method with the biasing fully turned on *ie* with $\lambda_B = 1$ (see references [1–3] and Eq. (9)) resulted in a very slow convergence rate for the heat capacity of hot *ST*2 [8] water. This result, previous references in the literature that suggest at least the potential for this [5], and the foregoing considerations, prompted the present article.

What is attempted here, is to further our understanding of the fundamental basis on which both Metropolis and force-biased Monte Carlo rest. Specifically, we would like to understand how these algorithms influence convergence rates and the stochastic characteristics of the Markov chains that they generate.

The remainder of this article is organized into three sections. The Theory section presents the problem and describes how the one-step transition matrices and their convergence-related properties are obtained. In the Results section, the values of these properties and their implications vis-a-vis convergence are reported and discussed, as are the stochastic characteristics of the corresponding Markov chains. In the Discussion section the main findings of this work are briefly summarized and descriptively explained, and the question of their relevance to actual simulations on liquids is discussed.

THEORY

a. *Construction of the Basic One-Step Transition Matrix*

The elements of the transition matrix $\| p_{ij} \|$, defined on M states $i = 1, 2, 3, \ldots, M$, give the probability of the system's undergoing a transition from state "i" to state"j." The expressions for the p_{ij}'s are obtained from the condition of detailed balance, which requires that the number of systems in the ensemble making the transition i to j, be equal to the number making the reverse transition j to i. This condition ensures that any arbitrary initial distribution function will eventually converge to the limiting distribution [7]. Thus, if $\{f_i\}$ is the normalized limit (equilibrium) distribution of the Markov chain, detailed balance requires that:

$$f_i p_{ij} = f_j p_{ji}, \qquad i \neq j. \tag{1}$$

Writing p_{ij} as the product of an a priori, normalized transition probability p^*_{ij}, times the acceptance probability a_{ij}, gives

$$\frac{a_{ij}}{a_{ji}} = \frac{f_j}{f_i} \frac{p^*_{ji}}{p^*_{ij}}, \qquad i \neq j. \tag{2}$$

When the right-hand side of Eq. (2) is greater than 1, we set $a_{ij}=1$, so that

$$a_{ji}=\frac{f_i p_{ij}^*}{f_j p_{ji}^*}, \qquad i\neq j. \tag{3}$$

Otherwise, set $a_{ji}=1$, so that

$$a_{ij}=\frac{f_j p_{ji}^*}{f_i p_{ij}^*}, \qquad i\neq j. \tag{4}$$

Thus we get the condition:

$$p_{ij}=\begin{cases} p_{ji}^*\dfrac{f_j}{f_i}, & \dfrac{f_j p_{ji}^*}{f_i p_{ij}^*}\leqslant 1 \\ p_{ij}^*, & \dfrac{f_j p_{ji}^*}{f_i p_{ij}^*}>1 \end{cases} \qquad i\neq j. \tag{5}$$

The diagonal matrix elements are obtained from the stochastic requirement

$$p_{ii}=1-\sum_{j\neq i}^{M} p_{ij}. \tag{6}$$

We will deal with a canonical ensemble, so that

$$f_i=e^{-\beta E_i}\Big/\sum_{i=1}^{M} e^{-\beta E_i}, \tag{7}$$

where E_i is the configurational energy of the system in state i, $\beta=(kT)^{-1}$, $k=$ Boltzmann's constant, $T=$ absolute temperature. Force-biased Monte Carlo (FBMC) and Metropolis Monte Carlo (MMC) differ in their choice of p_{ij}^*. In its most general form, the only requirement on the MMC a priori transition matrix, is that it be symmetric (i.e., $p_{ij}^*=p_{ji}^*$, $i\neq j$). In the simplest form of MMC, which we will use here, this symmetry requirement is met by the use of a uniform distribution over an interval. For this form of MMC:

$$p_{ij}^*=\begin{cases} 0, & j \text{ not accessible from } i, \\ \dfrac{1}{M_A(i)} & \text{for all accessible } j\neq i. \end{cases} \tag{8}$$

In Eq. (8) $M_A(i)+1$ is the number of accessible states given "i." In practice $M_A(i)\ll M$.

In FBMC the a priori transition matrix elements are obtained from

$$p_{ij}^*=\begin{cases} 0, & j \text{ not accessible from } i \\ \dfrac{e^{-\lambda_B\beta E_j}}{\sum_{j\neq i}^{M_A(i)+1} e^{-\lambda_B\beta E_j}} & \text{for all accessible } j\neq i, \end{cases} \tag{9}$$

where λ_B is an adjustable constant between 0 and 1, the subscript B designating "Berne." When $\lambda_B = 0$ we recover equation 8. Physically, p_{ij}^* for $\lambda_B > 0$ acts as a biasing function that causes the trial particle to preferentially seek out low energy states. Hereafter we will use the abbreviation $\lambda_B = \alpha$ to mean "the FBMC algorithm with λ_B set equal to the value α." $\lambda_B = 0$ will mean "the MMC algorithm."

b. *Scaling Down the Number of States*

Both the total number of system states M, and the number of system states accessible from one of them, $M_A(i) + 1$, are astronomically large for a many-body fluid. Therefore it is impossible to do numerical work with the actual system transition matrix. It is, however, conceivable that the MMC and FBMC algorithms have convergence-related characteristics that are intrinsic to the algorithm, i.e., that they display some generality and some degree of independence of the system and of the number of system states. Therefore it should be interesting, and perhaps ultimately profitable, to study the convergence characteristics of MMC and FBMC transition matrices of tractable size, which have some of the characteristics of the true underlying transition matrices. It was in this spirit that this work was undertaken.

It is clearly imperative to reduce the actual semi-infinite number of states to some manageable number, and we have done this by going over from a continuous to a discrete energy distribution function. The actual energy distribution functions of liquids are roughly Gaussian (see, e.g., [2, Fig. 2] for that of an *ST*2 fluid). Therefore, we approximate these essentially continuous functions by Gauss-like histograms. Thus we replace the essentially infinite number of energy levels available to the system by a small finite number of energy levels. It will be convenient to deal with single-particle energy distribution functions rather than the distribution functions for the mean energy. The former will have the same shape as the latter (roughly Gaussian) but the standard dviation of the former will exceed that of the latter by a factor of $\sqrt{N}$, where N is the number of particles [6, 9].

The number of system states in each level, i.e. the system degeneracy, is determined by the Gaussian nature of the energy distribution function and the Boltzmann nature of the state distribution function. Thus the relative system degeneracies $\Omega(E)$, of two system levels "i" and "j" with reduced energies βE_i and βE_j are obtained as follows:

$$\frac{f_l(i)}{f_l(j)} = \frac{\exp(-(\beta\bar{E} - \beta E_i)^2/2\sigma^2)}{\exp(-(\beta\bar{E} - \beta E_j)^2/2\sigma^2)} = \frac{\Omega(E_i)\, e^{-\beta E_i}}{\Omega(E_j)\, e^{-\beta E_j}}, \tag{10}$$

where $\bar{E}$ and σ are the assumed mean and standard deviation of the continuous single-particle distribution function, and $f_l(i)/f_l(j)$ is the ratio of probabilities that a randomly selected particle will be in levels "i" versus "j." Also $f_l(i)/f_l(j)$ is the ratio of systems in the ensemble in levels "i" and "j." Of course this procedure allows us to obtain only relative degeneracies; to obtain absolute values of $\Omega(E)$ we take the least degenerate (most stable) level to be singly degenerate. The number of system

states M on which our one-step matrix is defined, i.e., the order of the matrix is therefore given by

$$M = \sum_{k=1}^{L} \Omega_k(E), \qquad L = \text{number of levels.} \tag{11}$$

Depending on the application, we will deal with a variety of matrices ranging in size up to 276×276.

c. *Building in Operational Constraints*

(i) *Limited accessibility.* As mentioned previously, the number of system states accessible at each step, $M_A(i)+1$, is much less than the total number of system states M. While we cannot, because of size limitations, realisticaly incorporate this feature, we can develop some understanding of the role of limited accessibility by doing calculations on a series of matrices for which the fraction of accessible states varies.

Suppose, therefore, that up to $(2N+1)$ states (up to N states above "i" and up to N states below "i"), are accessible from each of the M states of our one-step matrix. Then the modified matrix elements for $0 \leqslant \lambda_B \leqslant 1$ are obtained from

$$p_{ij}^* = \begin{cases} e^{-\lambda_B \beta E_j} \Big/ \sum_{j=i-N}^{j=i+N}{}' e^{-\lambda_B \beta E_j}, & i \neq j, \quad |i-j| \leqslant N \\ 0, & i \neq j, \quad |i-j| > N, \end{cases} \tag{12}$$

where the prime indicates $i \neq j$ and truncation of the sum if j goes below 1 or above M. The modified values of p_{ij} and p_{ii} are obtained by using Eq. (12) in Eqs. (5) and (6). Equation (12) means that up to N neighbouring states with indices above, and up to N neighbouring states with indices below each state, are accessible at each step. Our states will be numbered sequencially with energy (see Results (A), (C)) so that in cases of limited accessibility, the accessibility is to degenerate and to energetically neighbouring states. The constraint implicit in Eq. (12), that when j is inaccessible from i in one step then i must be inaccessible from j in one step is required in order to preserve detailed balance.

(ii) *Distortion in the force-bias.* To use Eq. (12) with $\lambda_B > 0$ we must have values for E_j, for all accessible $j \neq i$, when system is in state "i." This is the main problem associated with FBMC. While it is possible in principle to know this, by taking a large number of successive gradients of the potential of the trial particle in state "i," and then Taylor expanding to high order around state "i," such a calculation would be prohibitively time-consuming, especially in a Monte Carlo simulation. What is done in practice, is to take one gradient of the potential, i.e., to obtain the forces (and torques if nonzero) on the trial particle in state "i" and then to estimate E_j, at all accessible j's through a first-order Taylor expansion around E_i. Therefore, the p_{ij}^* values used in practice (for $\lambda_B > 0$) will differ from those obtained by Eq. (12), and in this sense will be distorted.

We want to determine the consequences of working with distorted values of the p_{ij}^*'s for $\lambda_B > 0$, on both the convergence properties of the transition matrix, and on the stochastic characteristics of the Markov chains generated by the matrix. To do this we have to allow for the fact that the estimated values of E_j used to obtain p_{ij}^* in practice, will differ from the actual E_j values required in Eq. (12).

Here again, we are forced to resort to a rough approximation of the real situation. This is because these distortions depend on E_i, on the $(E_i - E_j)$'s, the state conditions, the maximum move size, and the force law all in a complicated and largely unknown way. To get some feeling for what is involved, consider the complexity that arises as a consequence of different maximum allowed displacements. Small maximum displacements will usually involve E_j estimates that are too low, since particles tend to be close to local minima where omission of the gradient of the force relative to the (small) force would produce an underestimate. However, larger maximum displacements can result in E_j estimates that are both low and high. High estimates result when the trial state involves a second physically neighbouring local minimum whose energy is below that of the original state. For example, a moderately large rotation of a water-like molecule will sometimes involve this latter phenomenon.

Also our use of histograms rather than continuous distribution functions would create unphysical discontinuities in our results if we tried, for example, to bring $(E_i - E_j)$ into the estimation of the error in E_j.

In view of all this, and because we are after only the main trends induced by the use of distorted p_{ij}^*'s we simplify the situation by writing:

$$E_j = E_j'(1 + \xi \Delta), \qquad j = 1, M, \tag{13}$$

where ξ is a random number between -1 and $+1$, Δ is a parameter that controls the average uncertainty associated with our estimate of E_j, and E_j' is the actual energy of state j, which in the present artificial situation is known. Thus $\Delta = 0$ gives no error, $\Delta = 1$ gives an error that can be as large as 100% in either direction around the true value, etc. When using Eq. (13), to obtain the E_j values for Eq. (12), we applied it with ξ reselected anew (uniformly between -1 and $+1$), once for each nonzero p_{ij}^* element. Thus a systematic variation of Δ, for fixed λ_B and fixed accessibility N, provides a means of studying the influence of using a distorted bias on the convergence and stochastic properties that will be obtained.

d. *Calculation of Eigenvalues and a Convergence-Related Standard Deviation of the Matrices*

Nine years ago Valleau and Whittington in an interesting article [10], drew attention to a property of the transition matrix, that can be used as a measure of the rate of convergence of a Markov chain to the limit distribution. For purposes of completeness we outline the relevant section of their work here.

If $f_i(t)$ is the probability that the system is in state "i" after t steps in a Markov chain, then

$$f_j(t+1) = \sum_{i=1}^{M} p_{ij} f_i(t) \tag{14}$$

and the solution of this set of difference equations, can be shown by substitution to be

$$f_j(t) = \sum_{k=1}^{M} \alpha_{jk} \lambda_k^t, \tag{15}$$

where λ_k are the eigenvalues of $\|p_{ij}\|$ and α_{jk} are constants related to the left eigenvectors of $\|p_{ij}\|$ and to the initial conditions. Since $\|p_{ij}\|$ is stochastic it will have one unit eigenvalue and all other eigenvalues will have modulii less than unity [11], provided the chain is irreducible, i.e. provided every state is accessible, within a finte number of steps, from every other state. Therefore in a long realization the convergence rate will be determined by that nonunit eigenvalue whose modulus is closest to unity. From Eq. (15) we see that small values of this eigenvalue correspond to rapid convergence rates. We adopt Valleau and Whittington's designation, the "subdominant eigenvalue" for this quantity, and give it their symbol λ_{sub}. λ_{sub} should not be confused with the FBMC parameter λ_B which is defined with Eq. (9).

Obviously, the reason that this property has previously been overlooked in comparing the relative efficiencies of, for example, FBMC and Smart Monte Carlo [3] is that one cannot extract the subdominant eigenvalue from a matrix defined over a semi-infinite number of states. But our matrices, which are based on the states in our Gauss-like histograms, will be defined over only tens or hundreds of states. For these latter dimensions numerical methods exist for obtaining good approximations for the eigenvalues of a matrix [12–14].

All our transition matrices were real and nonsymmetric. The procedure to get the eigenvalues involved transformation of the matrix to a real upper Hessenberg form followed by back-transformation to create the eigenvectors of the original matrix. The entire algorithm is available as a canned package in the IMSL library of most main frame computers [15]. The subdominant eigenvalues reported below were obtained with this algorithm. The accuracy of the values of λ_{sub} is, for large matrices, difficult to assess. The accuracy deteriorates with increasing size and increasing degeneracy of the matrix. Therefore all our calculations were done in double precision (15 figures). Also we report λ_{sub} values only for those matrices in which all the M eigenvalues converged, upon successive iterations, to within a preset standard. We believe that all our reported values of λ_{sub} are sufficiently accurate for our purpose, which is to compare relative values of λ_{sub} and to see how λ_{sub} is influenced by changes in our variables.

Another convergence-related property of the transition matrix (also proposed in

[10]), which has the value of being easier to obtain than λ_{sub}, is the standard deviation[1]:

$$\sigma = \sum_{i=1}^{M} f_i \sum_{j=1}^{M} |p_{ji} - f_i|. \tag{16}$$

The idea here is that if $p_{ji} = f_i$ for all j, then we will reach the limit distribution from any arbitrary initial distribution in one step. Therefore small values of σ should indicate rapid convergence rates and σ should correlate positively with λ_{sub}.

Results

a. *The Effect of the Number of States*

As previously mentioned, we bring the number of states down to a manageable number by discretizing the underlying energy distribution function, allotting one state to the least degenerate of the discretized levels, and then determining the remaining degeneracies by Eq. (10) and the total number of states M by Eq. (11). In this section we show that the use of a small number of states to represent the underlying distribution function does not seriously affect the calculated values of λ_{sub} or σ of the matrix.

We start by considering the very simple system: a two-level, M-state system, in which: $\beta E_1 = 1$, $\beta E_2 = 2$; $\Omega_1 = \Omega_2 = M/2$; and the states are distributed as a Boltzmann distribution, i.e., $f_l(1)/f_l(2) = e$. We examine the influence of M and Ω on the values of λ_{sub} and σ corresponding to the underlying one-step transition matrix, with $\lambda_B = 0$.

For $M = 2$, the transition matrix obtained from Eqs. (5)–(8), for this distribution function is

$$\mathbf{P} = \begin{pmatrix} 1 - e^{-1} & e^{-1} \\ 1 & 0 \end{pmatrix}$$

for which the eigenvalues are $(1, -e^{-1})$, so that $|\lambda_{sub}| = e^{-1}$. Using the numerical methods previously discussed, we have repeated the evaluation of λ_{sub} for $M = 4$, 8, 16, 32, 64, 128, and we have obtained σ (by Eq. (16)) for $M = 2$, 4, 8, 16, 32, 64, 128, 256, 512. (Our numerical procedure for evaluating λ_{sub} did not converge in this example for $M > 128$). In this way we determine the influence of the number of states assigned to each discretized level in our histograms on the values of λ_{sub} and σ of the corresponding one-step transition matrix. Of course, the mean and standard deviation of the distribution function are invariant with M because we scale Ω by the same factor in each level. So what we are doing here is seeing how, for an

[1] Our σ differs from the one in [10] by a factor of M. We have included M in our σ since we will want to compare the σ values of matrices of different orders.

otherwise fixed distribution function, the number of states used to describe the distribution function influences the values of λ_{sub} and σ of the corresponding transition matrix.

The results are displayed in Fig. 1, from which it is seen that both $|\lambda_{sub}|$ and σ have very similar forms (they are equal for $M=2$), they asymptotically approach a limiting value, and for $M=32$ we are already within $\sim 10\%$ of the asymptotic value of either λ_{sub} or σ. Also the values of λ_{sub} and σ become monotonic in M, provided M significantly exceeds the number of levels (here two). This is our first semi-rigorous indication that a relatively small number of states should be adequate for our purposes.

We reinforce this point by doing a calculation similar to the one above, but applied to a distribution function that more closely resembles those typical of liquids. Consider now a 9-level Gauss-like histogram distribution function, for which the relative probability of finding a particle in a level of energy βE_i is

$$f_i(\beta E_i) = e^{-(\beta E_i - u)^2/2w^2}. \tag{17}$$

Suppose $\beta E_1 = 1$, $\beta E_9 = 3$, and $\beta(E_{i+1} - E_i) = 0.25$. Then a choice of $u=2$ and $w=1$ in Eq. (17), together with $\Omega_1 = 1$ and the application of Eqs. (10) and (11)

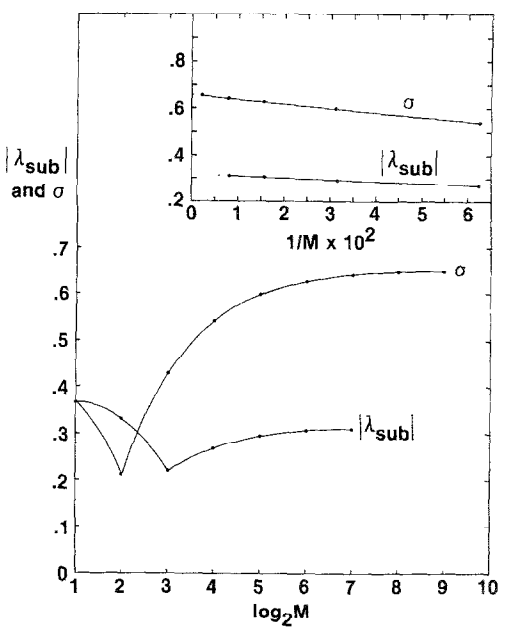

FIG. 1. Variation of the subdominant eigenvalue λ_{sub} and the standard deviation σ, with number of states M, for a one-step transition matrix based on a Metropolis algorithm and a 2-level, M-state distribution function with an equal degeneracy of $M/2$ in each level. λ_{sub} is defined under Eq. (15); σ is defined by Eq. (16). The lines and curves were drawn by eye through the points, both here and in all the other figures.

TABLE I

Values of λ_{sub} and σ at various N's and M's for One-Step Transition Matrices for the FBMC Algorithm with $\lambda_B = \frac{1}{2}, 1$[a]

		$\lambda_B = \frac{1}{2}$						$\lambda_B = 1$					
		$N = M/3$		$N = 2M/3$		$N = M$		$N = M/3$		$N = 2M/3$		$N = M$	
$\Omega(E_1)$[b]	M	λ_{sub}[c]	σ[d]	λ_{sub}[c]	σ[d]	λ_{sub}[c]	σ[d]	λ_{sub}[c]	σ[d]	λ_{sub}[c]	σ[d]	λ_{sub}[c]	σ[d]
1	39	0.8306	1.1533	0.5145	0.7024	0.4013	0.4943	0.8525	1.1545	0.5415	0.3881	0.0366	0.05446
2	78	0.8391	1.1678	0.5256	0.7251	0.4101	0.5190	0.8599	1.1639	0.5535	0.3865	0.0187	0.02775
3	117	0.8419	1.1729	0.5292	0.7328	0.4130	0.5273	0.8623	1.1676	0.5575	0.3866	0.0124	0.01861
4	156	0.8433	1.1756	0.5311	0.7370	—	0.5315	0.8635	1.1697	0.5594	0.3869	—	0.01400
5	195	0.8441	1.1772	0.5322	0.7396	—	0.5340	0.8642	1.1709	0.5606	0.3870	—	0.01122
6	234	0.8447	1.1784	—	0.7413	—	0.5357	0.8647	1.1718	0.5613	0.3872	—	0.00936
7	273	—	1.1792	—	0.7425	—	0.5369	0.8651	1.1724	—	0.3873	—	0.00803
8.	312	—	1.1789	—	0.7434	—	0.5378	—	1.1729	—	0.3874	—	0.00703
9	351	—	1.1802	—	0.7441	—	0.5385	—	1.1733	—	0.3874	—	0.00625
10	390	—	1.1806	—	0.7447	—	0.5390	—	1.1736	—	0.3875	—	0.00563

Note. In all cases $\Delta = 0$.[a] The matrices are defined over the states in a 9-level M-state Gauss-like histogram distribution function of mean 1.9952 and standard deviation 0.60273 (see text).

[a] These quantities are used in Eqs. (12) and (13).

[b] $\Omega(E_1)$ is the number of states allotted to the least degenerate (lowest energy) level.

[c] Where entries are not given, we were unable to obtain convergent solutions for λ_{sub}.

[d] Calculated by Eq. (16).

results in the 39-state Gauss-like histogram distribution function with the degeneracies: $\Omega_1 = 1$, $\Omega_2 = 2$, $\Omega_3 = 2$, $\Omega_4 = 3$, $\Omega_5 = 4$, $\Omega_6 = 6$, $\Omega_7 = 7$, $\Omega_8 = 7$, $\Omega_9 = 7$. This distribution function is liquid-like in that its standard deviation (0.60273) is about 30% of its mean (1.9952). (For common liquids at ordinary state conditions single-particle distribution functions have standard deviations of $\sim 20\%$–50% of the mean; see text, Theory, part b and [6]).

Again, we are interested in seeing how increasing the number of states, by proportionately scaling the number allocated to each level, will influence the values of λ_{sub}. Again, we do not change the mean or standard deviation of our distribution function by this scaling procedure. We have done this calculation for $\lambda_B = \frac{1}{2}$, and 1 and $N = M/3$, $2M/3$, and M in Eq. (12), and $\Delta = 0$ in Eq. (13) (i.e., no distortion in the p_{ij}^*'s). The $\lambda_B > 0$ choice was dictated by the need for our numerical procedure to converge with increasing numbers of states; the choice $\Delta = 0$ was made to ensure smoothness of the values of λ_{sub} with increasing numbers of states. Different values of λ_B and N were used to study the M-dependence of the value of λ_{sub} when different algorithms and different degrees of accessibility are used. It is seen from the entries in Table I and the lines in Fig. 2 that the values of λ_{sub} change very gradually with increasing numbers of states. Also, except for the $N = M$, $\lambda_B = 1$ calculation (which has the least connection with real simulations) the slopes for the λ_{sub} versus $1/M$ plots are all similar to one another, and in all cases (except $N = M$,

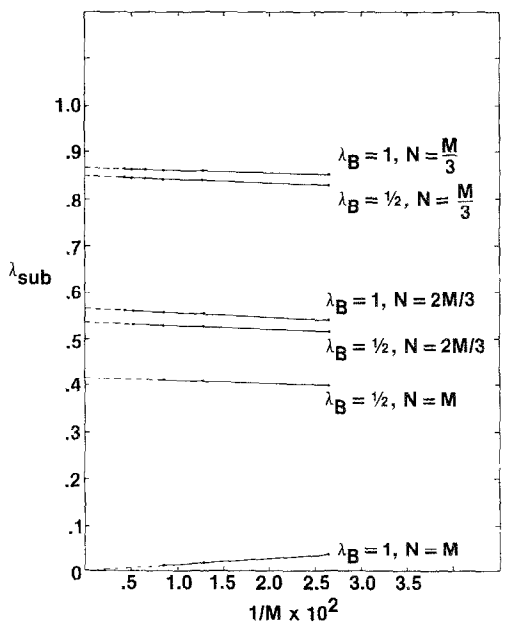

FIG. 2. Graphical display of the asymptotic behavior of λ_{sub} with M, for several values of N and λ_B, for transition matrices based on a 9-level Gauss-like histogram distribution function. λ_B, a biasing parameter, is defined under Eq. (9). Additional details are given with Table I.

$\lambda_B = 1$) we see that with only 39 states λ_{sub} is within $\sim 5\%$ of the appropriate asymptotic value.

In the calculations described in sections c–e below, the number of levels will be 9 and the number of states will be 276. The foregoing results support the contention that our calculations should have some bearing on the semi-infinite system.

b. *Correlation between λ_{sub} and σ*

In their study on transition matrices defined over relatively small numbers of states, Valleau and Whittington [10] pointed out that a close correspondence existed between λ_{sub} and σ. Since σ is trivial to calculate relative to λ_{sub}, we thought it worthwhile to further explore the extent of this correlation for the much larger matrices considered here. We already have the result from Fig. 1 that these two functions have strikingly similar functional forms. To examine the correlation in more detail we use the 9-level 39-state Gauss-like distribution function introduced in the previous section, to construct a series of transition matrices, using neighbour accessibility function N. The results are shown in Fig. 3. As expected, we

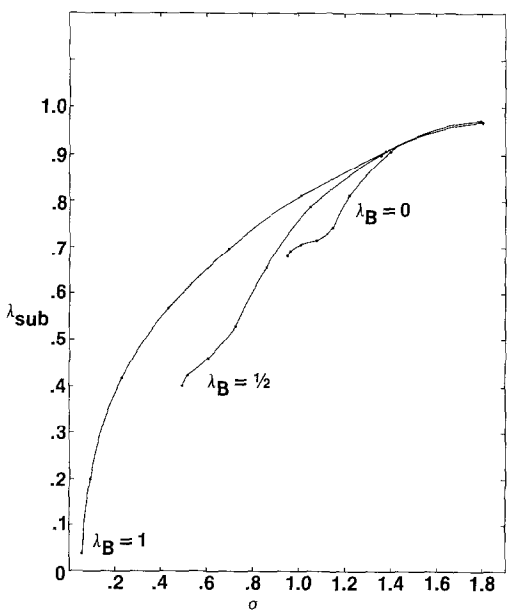

FIG. 3. The correlation between λ_{sub} and σ for a series of transition matrices defined over a 9-level 39-state Gauss-like histogram distribution function. In all cases Δ (in Eq. (13)) was zero. In each curve the left-most extreme point was for complete state-to-state accessibility, i.e., $N = 39$, and the rightmost extreme point was for the smallest accessibility considered, $N = 5$. Intermediate points are for $N = 35$, 30, 25, 20, 15, 10, respectively, going from left to right.

see that λ_{sub} and σ go down (indicating more rapid convegence) as the state-to-state accessibility goes up. We also see that the correlation between λ_{sub} and σ is fair for a fixed λ_B but it breaks down when λ_B is allowed to vary. This is seen from the fact that we get a family of curves and not one curve as λ_B is allowed to vary over the range 0–1. This implies that σ can be used in place of λ_{sub} as a rough measure of the convergence rate of a particular algorithm, but it cannot be used in place of λ_{sub} as a criterion for comparing the relative convergence rates of different algorithms. This is why we have used only λ_{sub} as a criterion for convergence rates in sections c and e below.

It is easy to show why $\lambda_B = 1$ for complete accessibility and distortion-free p_{ij}^*'s gave the very low values of σ and λ_{sub} shown in the two right-most columns of Table 1 and at the left of the $\lambda_B = 1$ curve in Fig. 3. Substituting Eqs. (7) and (9) into Eq. (5), and setting $\lambda_B = 1$ gives

$$p_{ij} = \begin{cases} e^{-\beta E_j} \Big/ \sum_{i \neq j}^{M} e^{-\beta E_i}, & \dfrac{f_j p_{ji}^*}{f_i p_{ij}^*} \leqslant 1 \\ e^{-\beta E_j} \Big/ \sum_{j \neq i}^{M} e^{-\beta E_j}, & \dfrac{f_j p_{ji}^*}{f_i p_{ij}^*} > 1 \end{cases} \qquad i \neq j.$$

Since both denominators in the expressions for p_{ij} will be very close to $\sum_{i=1}^{M} e^{-\beta E_i}$ for sufficiently large number of states, we get the result that for large numbers of states p_{ij} approaches f_j for all i. That is, for this choice of parameters the one-step transition matrix approaches the infinite-step transition matrix. Also σ, by Eq. (16), approaches zero for this matrix. Thus, if Eqs. (5) and (9) apply, i.e., if p_{ij}^* is known without distortion, and every state is at each step accessible from every other state, our one-step transition matrix provides convergence to the limit distribution, from any arbitrary starting distribution, in a single step. It will become clear in the next section, that violating these conditions on p_{ij}^* or on the accessibility, destroys this ideal situation.

c. *The Influence of Accessibility, λ_B, and distortion in p_{ij}^*, on λ_{sub}*

We now examine how limited state-to-state accessibility (controlled by N in Eq. (12)), distortion in p_{ij}^* for $\lambda_B > 0$ (controlled by Δ in Eq. (13)), and different degrees of biasing (controlled by λ_B in Eq. (12)) affect the value of λ_{sub} of our one-step transition matrices. We are particularly interested in the possible bearing of these results to simulations on dense liquids at ordinary state conditions, so our histogram distribution function was defined over the 9-levels: $\beta E_1 = -7$ to $\beta E_9 = -3$ with $\beta(E_{i+1} - E_i) = 0.50$, and with u and w in Eq. (17) equal to -5.0 and 1.2, respectively. Then Eqs. (10) and (17) generate the 9-level, 276-state, Gauss-like limit distribution with a mean $\beta\bar{E} = -4.9998$, and a standard deviation $= 1.0237$. From Eq. (10) the states are distributed over the levels as: $\Omega_1 = 1$, $\Omega_2 = 3$, $\Omega_3 = 8$, $\Omega_4 = 16$, $\Omega_5 = 30$, $\Omega_6 = 45$, $\Omega_7 = 57$, $\Omega_8 = 61$, $\Omega_9 = 55$. This single-particle distribution function resembles that for dense liquids in that the standard deviation is

about 20% of the mean. The parameters were chosen also because with this choice, M was 276: a number which is small enough so that our numerical determination of λ_{sub} was feasible (in most cases), and large enough to provide values of λ_{sub} that should be close to the asymptotic limit (see Section a and Fig. 2). This distribution was used with Eqs. (5)–(7), (12), and (13) to generate all the transition matrices used in this section.

Our results are displayed in Fig. 4 as cross-sectional slices through the families of 3-dimensional surfaces that arise. All the values of λ_{sub} reported in this section were real and positive.

Figure 4a show how N and Δ influence λ_{sub} at a fixed value of λ_B ($=\frac{1}{2}$). It is seen that provided N is low, λ_{sub} decreases, indicating more rapid convergence, as the number of accessible states available in one step increases, and this effect is greater, the smaller is Δ, i.e., the smaller the distortion in p_{ij}^*. This is obviously reasonable.

Figure 4b shows how Δ and λ_B influence λ_{sub} at fixed limited accessibility N ($=100$). The minima displayed by these graphs in the region $0<\lambda_B<1$ are the first theoretically based indication that in an actual realization, FBMC may be expected to converge at an optimal rate if the value of λ_B is taken to be between 0 and 1. Clearly, for all the values of Δ studied, we see that there exist values of λ_B between 0 and 1, that give rise to smaller values of λ_{sub} than are obtained with

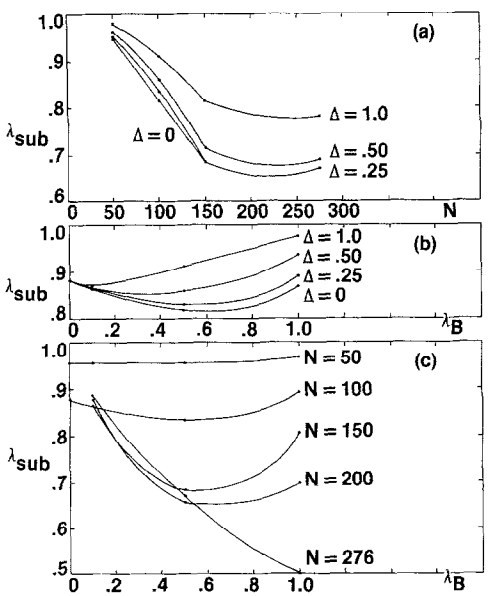

FIG. 4. The results in Figs. a–c are for transition matrices defined over the 9-level 276-state Gauss-like histogram distribution function described in part c of the Results section. Figure a shows the effect of N and Δ on λ_{sub} at fixed λ_B ($=\frac{1}{2}$). Figure b shows the effect of Δ and λ_B on λ_{sub} at fixed N ($=100$). Figure c shows the effect of N and λ_B on λ_{sub} at fixed Δ ($=0.25$). Missing sections in Figs. (a) and (c) are due to nonconvergence of the algorithm used to get the eigenvalues and one point ($N=150$ $\lambda_B=0$ in Fig. 4c) was omitted because of a suspected inaccuracy.

either $\lambda_B = 0$ or 1. As expected, smaller values of Δ (i.e., less distorted p_{ij}^*'s) result in smaller λ_{sub}'s. Also, the smaller the value of Δ, the more is the minimum in λ_{sub} shifted toward higher values of λ_B. Again, this makes sense, because we would expect to make better use of FBMC, the less distorted is the biasing function p_{ij}^*. Note however, from Fig. 4b, that even with $\Delta = 0$, i.e., even with p_{ij}^* known without distortion, $\lambda_B = \frac{1}{2}$ yields a lower value of λ_{sub} than does $\lambda_B = 1$. Thus Fig. 4b demonstrates that a reduction of state-to-state accessibility from full accessibility (i.e., a reduction of N from 276 to 100) is alone sufficient to make $\lambda_B = \frac{1}{2}$ more efficient than $\lambda_B = 1$.

The results in Figs. 4a and b provide the explanation of why the extremely low λ_{sub}'s (and σ's) found in Fig. 3 for $\lambda_B = 1$ and complete accessibility are not realizable in practice. These very low values of λ_{sub} are obtained *only* when there is complete state-to-state accessibility at each step, and when the p_{ij}^*'s are known without distortion. Figure 4b shows that when either or both of these conditions are not met, an intermediate value of λ_B between 0 and 1, provides a more rapidly convergent algorithm.

Figure 4c illustrates how N and λ_B influence λ_{sub} at fixed Δ ($= 0.25$). These results again show the minimum in λ_{sub} at values of λ_B between 0 and 1, provided the accessibility N is less than complete, i.e., provided $N < 276$. (Although impossible to see with the scale used, there is a minimum at $\lambda_B \sim 0.1$ for the $N = 50$ curve). They also show how the minima are shifted to greater λ_B the greater the accessibility N. This is consistent with the results in Figs. 4a and b with the foregoing discussion.

d. *Asymptotic Acceptance Rates*

As the acceptance rate is a closely monitored quantity in a real simulation, we thought it worthwhile to present some of the accpetance rates that arose in our model calculations.

The average acceptance rate $\bar{R}(t)$ at any step t in the Markov chain was obtained by

$$\bar{R}(t) = \sum_{i=1}^{M} f_i(t)\, \overline{AR}_i, \tag{18}$$

where $f_i(t)$ is the probability of being in state i at step t, and $\overline{AR}_i$ is the average acceptance rate for a move from state i to any other state j, that is accessible in one step. Thus, from Eqs. (2)–(4) $\overline{AR}_i$ is obtained from

$$\overline{AR}_i = \frac{\sum'_{j=i-N}^{j=i+N} a_{ij}}{\sum'_{j=i-N}^{j=i+N} 1}, \qquad i \neq j, \tag{19}$$

$$a_{ij} = \begin{cases} 1 & \text{if } \dfrac{f_j(\infty)}{f_i(\infty)} \dfrac{p_{ji}^*}{p_{ij}^*} > 1 \\ \dfrac{f_j(\infty)}{f_i(\infty)} \dfrac{p_{ji}^*}{p_{ij}^*} & \text{otherwise} \end{cases} \tag{20}$$

and the meaning of $\sum'$ is as given with Eq. (12). The asymptotic mean acceptance rates we report ($\bar{R}_\infty$) were obtained by using the limit distribution $\{f_i(\infty)\}$ for $f_i(t)$ in Eq. (18).

A sample of our results that illustrate the main characteristics are displayed in Figs. 5a and b. The results shown are for the 276-state, 9-level distribution function discussed in Section c.

Figure 5a shows how $\bar{R}_\infty$ varies with increasing accessibility of states N, for p_{ij}^* distortion-free, and for $\lambda_B = 0$, 0.5, and 1. As expected, with no distortion allowed in $\lambda_B = 0$ or $\frac{1}{2}$ $\bar{R}_\infty$ decreases significantly with increasing accessibility. The very high values of $\bar{R}_\infty$ for $\lambda_B = 1$ are clearly a consequence of disallowing distortion in p_{ij}^*. It is interesting to compare the $\bar{R}_\infty$ values for $\lambda_B = 1$ in Fig. 5a with the λ_{sub} values for $\lambda_B = 1$ in Fig. 4b. This comparison reveals that the high acceptance rates at limited accessibilities shown in Fig. 5a (e.g., $\bar{R}_\infty = 0.902$ at $N = 100$) do not correspond to particularly low values of λ_{sub} at these limited accessibilities ($\lambda_{sub} = 0.867$ for $\lambda_B = 1$, $\Delta = 0$, $N = 100$). So high acceptance rates need not go hand-in-hand with rapid convergence rates.

As seen from Fig. 5b the situation changes drastically when we allow for distortion in the p_{ij}^*'s. Here we plot $\bar{R}_\infty$ as a function of λ_B and Δ for a fixed value of the

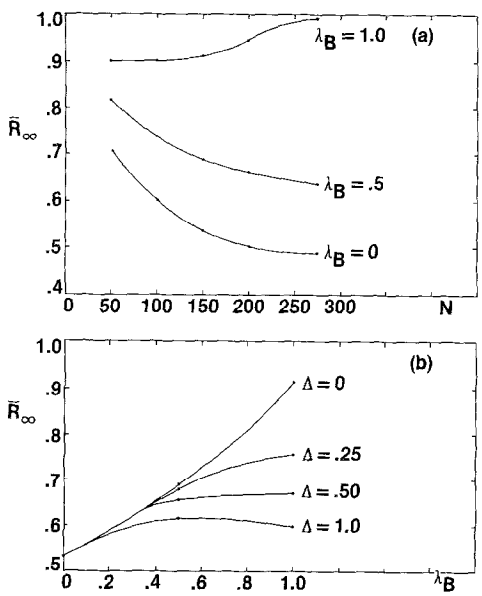

FIG. 5. The asymptotic acceptance rates ($\bar{R}_\infty$) for transition matrices defined over the distribution function described in the caption to Fig. 4 and part c of the results section. Figure a shows the dependence of $\bar{R}_\infty$ on N and λ_B for fixed Δ ($= 0$). Figure b shows its dependence on λ_B and Δ at fixed N ($= 150$).

accessibility function N ($= 150$). Again, as expected, increasing the distortion in p_{ij}^*, generally results in a drop in $\bar{R}_\infty$ and this effect becomes greater the larger λ_B.

e. *Stochastic Characteristics of the Markov Chains*

Here we examine the time evolution of the distribution functions that are generated by the various transition matrices. What we do is to monitor the stepwise development of the relevant characteristics of the generated distribution functions, starting from physically important initial distribution functions. At any time t, the new state distribution function $\{f(t)\}$ is generated by the operation

$$\{f(t)\} = \{f(t-1)\}\,\mathbf{P}, \tag{21}$$

where $\mathbf{P}$ is any of the one-step transition matrices under consideration, $\{f(t)\}$ is the row vector

$$\{f(t)\} = (f_1(t), f_2(t), \ldots, f_M(t)). \tag{22}$$

M, as before, is the number of states in the system and $f_i(t)$ is the normalized probability that the system is in state i after t steps. Since we deal with a canonical ensemble,

$$f_i(\infty) = e^{-\beta E_i} \Big/ \sum_{i=1}^{M} e^{-\beta E_i}. \tag{23}$$

To make contact with the empirical results from real simulations, we define the energy level distribution function

$$\{F(t)\} = (F_1(t), F_2(t), \ldots, F_k(t)), \tag{24}$$

where k is the number of energy levels in the system and $F_l(t)$ is the probability that the system is in any state of level l at time t. Thus, $F_l(t)$ is obtained from

$$F_l(t) = \sum_{\Omega_l} f_i(t), \tag{25}$$

where Ω_l, as before, is the degeneracy of level l. Equation (25) means we sum over those states "i" assigned to energy level l. Thus the average energy of the system and its standard deviation, at each step, are calculated from

$$\bar{E}(t) = \sum_{l=1}^{k} F_l(t)\, E_l, \tag{26}$$

$$\mathrm{SDE}(t) = \left[\sum_{l=1}^{k} F_l(t)(E_l - \bar{E}(t))^2 \right]^{1/2} \tag{27}$$

respectively. The quantity SDE(t) in Eq. (27) is proportional to the square root of the configurational constant volume heat capacity (Cv) [2, 6], and so we will use

SDE(t) as an index of Cv(t). Also, to learn about the time-dependence by which a particular transition matrix fills in the low- and high-energy wings of the energy distribution we will simply monitor the values of $F_l(t)$ for the lowest and highest energy levels in our distribution function. The relative rate with which the lowest and highest energy levels develop toward their infinite time values, provides us with an index of how a particular algorithm, over time, generates low- and high-energy configurations.

We also want to see how these stochastic characteristic correlate with the values of λ_{sub} of the generating matrices. This correlation is not at all obvious because λ_{sub} depends only on the transition matrix, while the distribution function at any finite time t depends on both the transition matrix and the initial distribution function. Therefore we picked three transition matrices from Section c for which values of λ_{sub} were calculable. Specifically all the generating matrices are defined over the same 9-level, 276-state distribution function used in Section c. The three matrices corresponded to $\lambda_B = 0$, $\frac{1}{2}$, and 1. The maximum number of accessible neighbouring states N at each step was between 50 and 150, and Δ was taken to be 0.25 (Δ has no effect when $\lambda_B = 0$).

It remains to select the starting distribution functions $\{f(0)\}$. We do this by referring to a typical situation that arises in a real simulation. To evaluate functions that come in slowly, such as the heat capacities or correlation functions, one usually

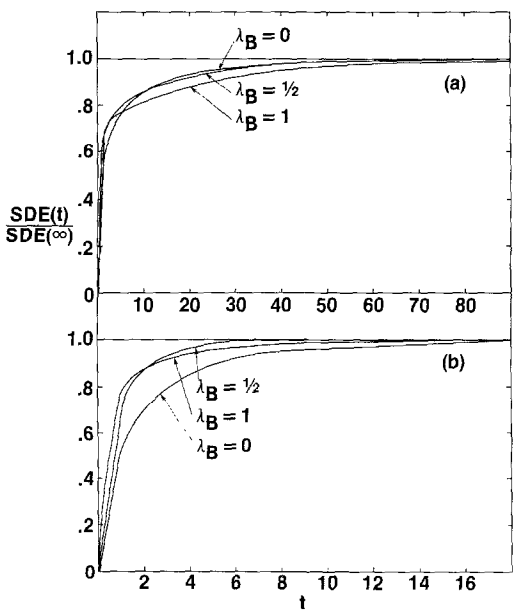

FIG. 6. The effect of λ_B on the rate of development of the standard deviation of the energy of the distribution function described in the caption to Fig. 4 and in part c of the Results; t is the number of steps in the Markov chain, $\Delta = 0.25$, and the initial distribution function is given by Eq. (28) for all the curves. SDE(∞) = 1.0237. Fig. 6a, $N = 50$; Fig. 6b, $N = 150$.

starts with a configuration that is in a sense typical of the limit distribution function, i.e., one that is energetically near the mean. In practice this is selected from one of the configurations generated after an initial equilibriation period. Thus, one of our initial distribution functions was

$$\{f_i(0)\} = \begin{cases} 0, & i = 1 \text{ to } 42, 44 \text{ to } 276 \\ 1, & i = 43. \end{cases} \tag{28}$$

The situation of having state 43 occupied was selected as a typical starting situation since, in our numbering scheme, 43 was in the middle of the 30 states assigned to level 5, whose energy was -5, which is close to the asymptotic mean of -4.9998. (See Results, Section c.)

We thought it interesting also to include in this study, chains that are generated when starting from an atypically high- and atypically low-energy configuration. So two other initial distribution functions we used were

$$\{f_i(0)\} = \begin{cases} 0, & i = 2 \text{ to } 276 \\ 1, & i = 1 \end{cases} \tag{29}$$

and

$$\{f_i(0)\} = \begin{cases} 1, & i = 276 \\ 0, & i = 1 \text{ to } 275, \end{cases} \tag{30}$$

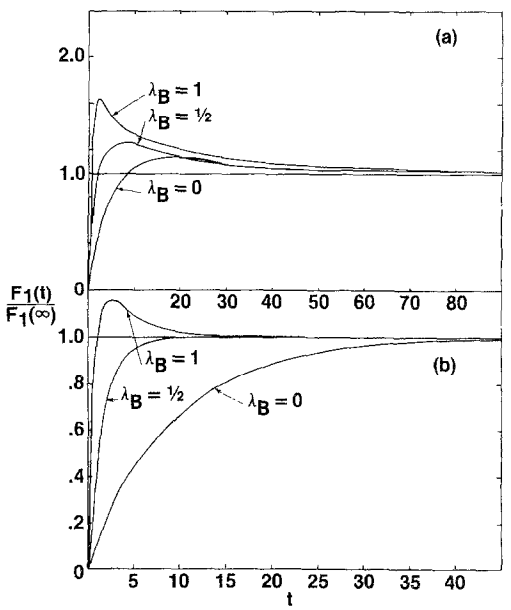

FIG. 7. The effect of λ_B on the rate of development of the lowest energy level in the distribution function. The conditions and limit distribution function are identical to those in caption to Figs. 6a and b: $F_1(\infty) = 0.04391$.

where we used Eqs. (29) and (30) to represent a low-energy and a high-energy initial configuration, respectively.

Our results are displayed in Figs. 6 to 8 for walks starting from a centrally located state (Eq. (28)) and in Fig. 9 for walks starting from our lowest and highest energy states (Eqs. (29) and (30)). The walks starting from a centrally located state were done for $N=50$ and 150 to demonstrate the influence of accessibility on the relative efficacy of the three algorithms. For purposes of clarity the plotted functions are all reduced with respect to their infinite time values. Therefore the "best" algorithm will be the one for which the plotted function approaches unity fastest.

We see from Fig. 6, that with respect to the rate of development of the standard deviation of the energy, the $\lambda_B=\frac{1}{2}$ algorithm is best in that it is either fastest (for $N=150$) or roughly tied for fastest (for $N=50$). Also, as the relative accessibility increases, we see that the $\lambda_B=1$ algorithm improves relative to the $\lambda_B=0$ algorithm and that the $\lambda_B=\frac{1}{2}$ algorithm improves relative to both the $\lambda_B=0$ and $\lambda_B=1$ algorithms. These results can be analyzed in terms of the results in Figs. 7 and 8 in which the rates of development of the lowest and the highest energy levels of our distribution function are displayed. It is seen from these figures that the $\lambda_B=1$ algorithm (for both $N=50$ and 150) fills in the lowest energy level from above (Figs. 6a and b) and the highest energy level from below (Figs. 7a and b) the

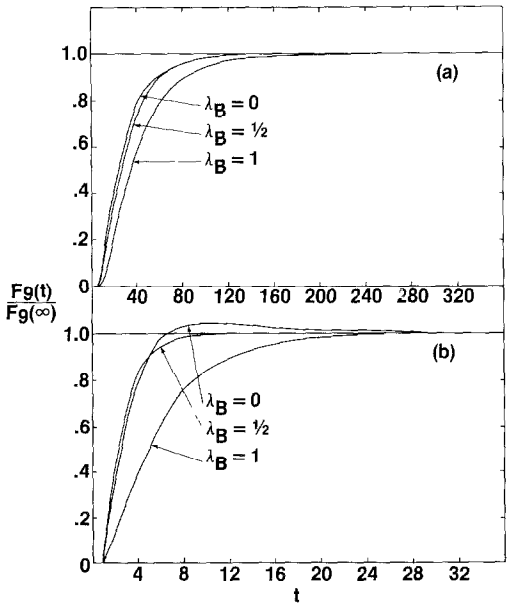

FIG. 8. The effect of λ_B on the rate of development of the highest energy level in the distribution function. The conditions and limit distribution function are identical to those in the caption to Figs. 6a and b, $F_9(\infty)=0.04423$.

asymptotic value. Also, the approach to these asymptots is slower for $\lambda_B = 1$ than for $\lambda_B = \frac{1}{2}$. Again the $\lambda_B = \frac{1}{2}$ algorithm displays the best overall performance in that it is either the fastest of the three (Figs. 7b, 8b) or essentially shares first place with [illegible] algorithm becomes slower relative to the $\lambda_B = \frac{1}{2}$ and $\lambda_B = 1$ algorithms as the accessibility increases.

An important feature of Figs. 6a, 7a, and 8a is their illustration of the tendency of the $\lambda_B = 1$ algorithm at relatively low accessibilities, to initially over-occupy low-energy levels at the expense of high-energy levels so as to produce an initial skewing of the distribution function toward its low-energy wing. This skewing results in the slow development of SDE(t) shown in Fig. 6a, since this latter quantity is positively correlated with the breadth of the energy distribution function.

Also, the results in Figs. 6–8 are consistent with the eigenvalue results shown in Fig. 4c in that each of the curves in Fig. 4c for $50 \leqslant N \leqslant 150$ and $\Delta = 0.25$ displayed a minimum for $0 < \lambda_B < 1$ indicating more rapid convergence for these intermediate λ_B values.

It is noteworthy that the relative superiority of the $\lambda_B = \frac{1}{2}$ algorithm occurs also in Figs. 9a and b where we show, respectively, the rate of development of our lowest level starting from a state in our highest level (our highest numbered state), and the

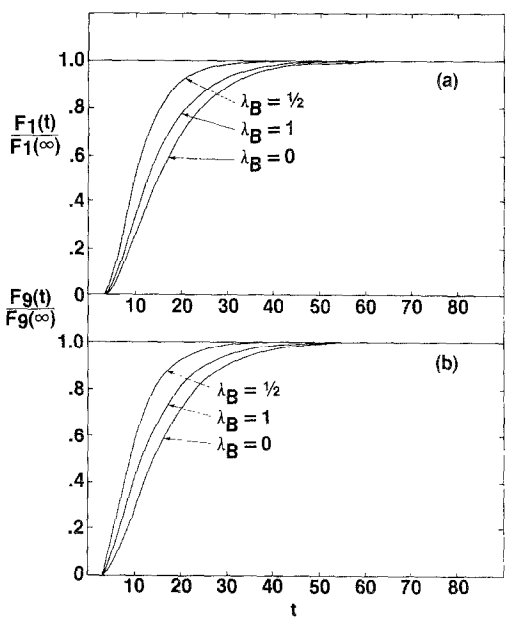

FIG. 9. (a) The effect of λ_B on the rate of development of F_1. The conditions and distribution function are identical to those in the caption to Fig. 6, except here the starting distribution function is that given by Eq. (30), and here $N = 100$. (b) The effect of λ_B on the rate of development of F_9. The conditions and distribution function are identical to those in the caption to Fig. 6, except here the starting distribution function is that given by Eq. (29), and $N = 100$.

development of our highest level starting from our lowest energy (and lowest numbered) state. The relative efficiencies are seen to be $(\lambda_B = \frac{1}{2}) > (\lambda_B = 1) > (\lambda_B = 0)$ for these walks.

Discussion

We have found both from the behavior of λ_{sub} of our model transition matrices, and from the stochastic characteristics of the distribution functions that are generated by these matrices, that FBMC with $0 < \lambda_B < 1$ on the whole provided better convergence characteristics than either the $\lambda_B = 0$ or the $\lambda_B = 1$ algorithms. The $\lambda_B = 1$ algorithm, in the initial stages of the walk, tended to produce an overoccupation of the low-energy levels at the expense of high-energy levels. This behavior was consistent with the slow heat capacity development that was found for the $\lambda_B = 1$ relative to the $\lambda_B = \frac{1}{2}$ algorithm.

It is worthwhile to note that an entirely different, previously published criterion, also leads to the conclusion that a choice of $\lambda_B = \frac{1}{2}$ is to be preferred [5]. This criterion is to pick λ_B so that the initial slope of the acceptance probability vs step-size curve is zero; this occurs with $\lambda_B = \frac{1}{2}$ and with "Smart Monte Carlo" but not with $\lambda_B = 0$ or 1.

The advantage entailed in using a value of λ_B intermediate between 0 and 1 at least for distribution functions of the kind adopted here[2] can be understood from the following qualitative argument. The $\lambda_B = 1$ algorithm is optimal and vastly superior to any other algorithm, but only when the biasing function p_{ij}^* is distortion-free and when there is complete state to state accessibility at each step. However, as soon as either of these conditions is violated, the use of $\lambda_B = 1$ becomes a form of over-biasing, in the sense that $\lambda_B = 1$ causes the search for low-energy states to be unjustifiably overzealous. On the other hand, some extra effort to seek out low-energy states is needed, because with the distribution functions we used, there are fewer low- than high-energy states available. Therein is the potential weakness in using $\lambda_B = 0$; by looking equally hard for low- and high-energy states, we occupy the low states too slowly, simply because without extra effort in the form of a bias they are not found frequently enough. Therefore, an intermediate value of λ_B (such as $\frac{1}{2}$) does best, because it provides some bias in the search for low-energy states, without pushing the biasing to a point justified only by ideal circumstances.

We are left with the important question of just how relevant these calculations and conclusions are to actual Monte Carlo simulations on dense molecular or atomic fluids. We have tried, by our choice of Gauss-like distribution functions and the introduction of limited accessibility and distortion in the p_{ij}^*'s, to make these calculations as relevant as possible. Nevertheless, the necessity to simplify through

[2] Of course our conclusions have little relevance to systems for which the energy distribution functions are grossly non-Gaussian. Strongly differing distribution functions for different systems, may well account for much of the observed [16] system-dependence of the efficiency of the different algorithms.

the use of histograms, and the approximate way in which distortion and accessibility was represented, all must contribute to a weakening of this connection.

We believe that these calculations are more relevant than not to actual simulations on atomic or molecular liquids. This view is based partially on our own [6] and partially on other workers' [16, 17] experience with FBMC simulations on such liquids. But this view must still be considered tentative since it does not come from a really systematic study. To put the matter to the test, we have run a series of simulations for the $\lambda_B = 0$, $\frac{1}{2}$, and 1 algorithms, designed to specifically look for our main predictions. An examination of these results is encouraging, as is seen from the article that immediately follows this one.

ACKNOWLEDGMENTS

We would like to thank Dr. Murray Alexander for his time and energy to help implement the IMSL routines, Dr. David Fieldhouse for his valuable comments on the M-dependence of the eigenvalue problem, and one of the referees for his perceptive and detailed review. We are also grateful to the Natural Sciences and Engineering Research Council of Canada for financial support.

REFERENCES

1. C. PANGALI, M. RAO, AND B. J. BERNE, *Chem. Phys. Letts.* **55** (1978), 413.
2. M. RAO, C. PANGALI, AND B. J. BERNE, *Mol. Phys.* **37** (1979), 1773.
3. M. RAO AND B. J. BERNE, *J. Chem. Phys.* **71** (1979), 129.
4. J. C. OWICKI AND H. A. SCHERAGA, *Chem. Phys. Letts.* **47** (1977), 600.
5. P. J. ROSSKY, J. D. DOLL, AND H. L. FRIEDMAN, *J. Chem. Phys.* **69** (1978), 4628.
6. S. GOLDMAN, *J. Chem. Phys.* **79** (1983), 3938.
7. N. METROPOLIS, A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, AND E. TELLER, *J. Chem. Phys.* **21** (1953), 1087.
8. F. H. STILLINGER AND A. RAHMAN, *J. Chem. Phys.* **60** (1974), 1545.
9. J. E. FREUND, "Mathematical Statistics," p. 177, Prentice–Hall, Englewood Cliffs, N.J., 1962.
10. J. P. VALLEAU AND S. G. WHITTINGTON, *J. Comput. Phys.* **24** (1977), 150.
11. E. F. BECKENBACH AND R. BELLMAN, "Inequalities," Springer-Verlag, Berlin, 1965.
12. J. H. WILKINSON, "The Algebraic Eigenvalue Problem," Oxford Univ. Press, (Clarendon) London/New York, 1965.
13. B. T. SMITH, J. M. BOYLE, B. S. GARBOW, Y. IKEBE, V. C. KLEMA, AND C. B. MOLER, "Matrix Eigensystem Routines," Springer-Verlag, Berlin, 1974.
14. A. S. HOUSEHOLDER, "The Theory of Matrices in Numerical Analysis," Blaisedell, New York, 1964.
15. "IMSL Library Reference Manual," 9th ed. IMSL, Houston, 1982.
16. Comments of an anonomous referee.
17. J. D. DOLL, "NRCC Proceedings No. 6," Conf. 790735, 52, 1979.