

# Expanded grand canonical and Gibbs ensemble Monte Carlo simulation of polymers

Fernando A. Escobedo and Juan J. de Pablo

Department of Chemical Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706-1691

(Received 17 April 1996; accepted 5 July 1996)

A novel formalism is presented for simulation of polymers in expanded grand canonical and expanded Gibbs ensembles. Molecular creation and destruction attempts are replaced by transition attempts between states of a tagged chain of variable length. Results are presented for expanded grand canonical simulations of hard-core chain fluids in the bulk and in a slit pore and for expanded Gibbs ensemble simulations of vapor-liquid equilibria for square-well chains. © 1996 American Institute of Physics. [S0021-9606(96)53134-X]

## I. INTRODUCTION

Insertion-type methods are ubiquitous in molecular simulations of phase equilibria.<sup>1-4</sup> Unfortunately, such methods tend to be inefficient for dense systems and for large, complex molecules. Development of more efficient insertion-type algorithms is a highly active area of research; many direct and indirect approaches have been proposed to permit simulation of systems for which older, conventional methods have failed.<sup>5-11</sup>

The idea of conducting a single simulation where different ensembles are probed (through a random walk over states) has been used before to facilitate the calculation of thermodynamic properties.<sup>12,13,8,9</sup> Attard,<sup>13</sup> Kaminsky,<sup>7</sup> and Wilding and Müller<sup>8</sup> have calculated the chemical potential of a fluid by gradual, isotropic coupling and decoupling of a tagged molecule from the rest of the system. Within the same expanded-ensemble framework, we recently proposed to introduce intermediate states of a tagged chain by varying its length (EVALENCH method).<sup>9</sup> The strength of the method resides in that the difficulty of inserting a whole molecule into the system at once (the probability of which is prohibitively low for large molecules and at intermediate to high densities) is reduced to that of inserting a few segments of the tagged chain.

In this Communication we develop expanded grand canonical (EGC) and expanded Gibbs (EG) ensemble simulations. The theoretical foundations are given in Sec. II; these are presented in a general form valid for arbitrary selection of the states of the tagged molecule. We specialize our prescriptions to states defined through a variable number of sites of the tagged chain. To illustrate the wide applicability of the methods proposed in this work, we present in Sec. III select simulation results of expanded GC simulations for hard chain molecules in the bulk and in a slitlike pore, and of expanded Gibbs ensemble simulations for square-well chain molecules. Some methodological aspects are discussed in Sec. IV.

## II. METHODOLOGY

### A. Expanded grand canonical ensemble

For a pure homopolymer system, an expanded ensemble partition function  $\Omega$  can be defined as follows:

$$\Omega = \sum_{n=0}^{\infty} \sum_{y=1}^M e^{N\beta\mu} e^{\Psi_y} Q(N, y, V, T), \quad (1)$$

where  $\mu$  is the chemical potential,  $\beta = (k_b T)^{-1}$  ( $k_b$  is Boltzmann's constant) and  $Q$  is the canonical partition function for a system of volume  $V$ , temperature  $T$ , and  $N$  (full) molecules and a tagged molecule in state  $y$ . States 1 and  $M$  correspond to a fully decoupled and a fully coupled molecule, respectively. The  $\Psi_y$ 's in Eq. (1) are (in principle) arbitrary preweighting factors whose relevance can be assessed as follows. The probability of encountering a particular configuration of the system (for given  $V$ ,  $T$ ,  $N$  and  $y$ ) is given by

$$p(y) = \frac{1}{\Omega} e^{N\beta\mu} e^{\Psi_y} e^{-\beta U(N, y)}, \quad (2)$$

where  $U(N, y)$  is the total energy of the system. By performing an ensemble average over all configurations at constant  $N$  and  $y$  we obtain

$$\langle p(y) \rangle = \frac{Q(N, y, V, T)}{\Omega} e^{N\beta\mu} e^{\Psi_y}. \quad (3)$$

To generate successful creations or destructions of the tagged molecule, the end states should be sampled as often as possible. This can be accomplished by selecting preweighting factors such that each state is visited with the same frequency. If for any two states  $x$  and  $y$  we prescribe that  $\langle p(x) \rangle / \langle p(y) \rangle = 1$ , it follows from Eq. (3) that

$$e^{\Psi_x - \Psi_y} \left[ \frac{Q(N, x, V, T)}{Q(N, y, V, T)} \right] = 1. \quad (4)$$

The logarithm of the term in brackets in Eq. (4) is proportional to the excess chemical potential associated with the transition  $x \rightarrow y$  of the system (under NVT conditions),<sup>9</sup> thus

$$\Psi_x - \Psi_y = -\beta \Delta \mu^E(x \rightarrow y) = \beta \mu^E(x, N) - \beta \mu^E(y, N). \quad (5)$$

Equation (5) suggests that a convenient choice for  $\Psi$  is

$$\Psi_y = \beta \mu^E(y, N) = \omega_y \mu^E(N) = \omega_y \left[ \beta \mu - \ln \left( \frac{N_y}{V} \right) \right], \quad (6)$$

where the definition of  $\mu^E(N)$  has been introduced (within an additive constant) and where  $N_y$  is the total number of molecules in the system, including the tagged chain if  $y \neq 1$ . The factors  $\omega_y$  to be used in (7) must satisfy  $\omega_1 = 0$  and  $\omega_M = 1$ ; intermediate values can be found from a preliminary run so that eventually each state is uniformly sampled. For homonuclear monodisperse polymer chains of  $n$  sites per chain, we choose to define the states by the number of sites in the tagged chain  $n_y (0 < n_y < n)$ . A convenient choice for  $\omega_y$  is provided by

$$\omega_y = \frac{n_y}{n}. \quad (7)$$

This prescription is based on the observation that the incremental chemical potential of a chain is relatively insensitive to chain length.<sup>14,15</sup>

A Monte Carlo simulation in the expanded ensemble is carried out in a manner analogous to that for a conventional grand canonical ensemble simulation. The difference lies in that molecular creation or destruction attempts are replaced by transition attempts between neighboring states of the tagged chain, i.e.,  $y \rightarrow y + 1$  or  $y \rightarrow y - 1$ . When the tagged chain reaches an end state, "another" chain becomes tagged. End states are actually indistinguishable from each other; a nominal state is assigned depending on whether an incremental or decremental transition is proposed (i.e.,  $y = 1$  for an incremental state,  $y = M$  otherwise). To facilitate transitions, the position of every appended bead to the tagged chain is chosen from a set of trial positions, favoring positions of low energy (continuum configurational-bias or CCB method<sup>16,17</sup>). Metropolis-type acceptance criteria are implemented for a transition from state  $y$  to state  $y + \Delta$ :<sup>9,16</sup>

$$P_{\text{accept}} = \min \left[ 1, \frac{K(y + \Delta \rightarrow y) p(y + \Delta)}{K(y \rightarrow y + \Delta) p(y)} \right] \\ = \min [1, (R_w)^\Delta \exp(\Psi_{y+\Delta} - \Psi_y)], \quad (8)$$

where  $R_w$  is the Rosenbluth weight associated with the growth process if  $\Delta = +1$  and with the hypothetical, reverse growth process if  $\Delta = -1$  ( $K$  is the probability of proposing a transition). It is given by

$$R_w = \prod_{i=n_A}^{n_B} N_{sp}^{-1}(i) \sum_{j=1}^{N_{sp}(i)} \exp[-\beta U_j(i)], \quad (9)$$

where  $n_A = \min(n_y, n_{y+\Delta}) + 1$  and  $n_B = \max(n_y, n_{y+\Delta})$ ,  $N_{sp}(i)$  is the number of random trial positions for the  $i$ th segment of the molecule, and  $U_j(i)$  is the interaction energy between the  $i$ th segment and the system.<sup>9,16</sup>

## B. Expanded Gibbs ensemble

The above formalism can be readily extended to simulations in the Gibbs ensemble.<sup>2</sup> The equilibration of chemical potential between the two simulation boxes is now achieved by attempting correlated transitions between states of a tagged molecule that is present in both boxes. Figure 1 illustrates the correlation between states of a tagged chain of  $n = 6$  beads for a particular choice of intermediate states; an

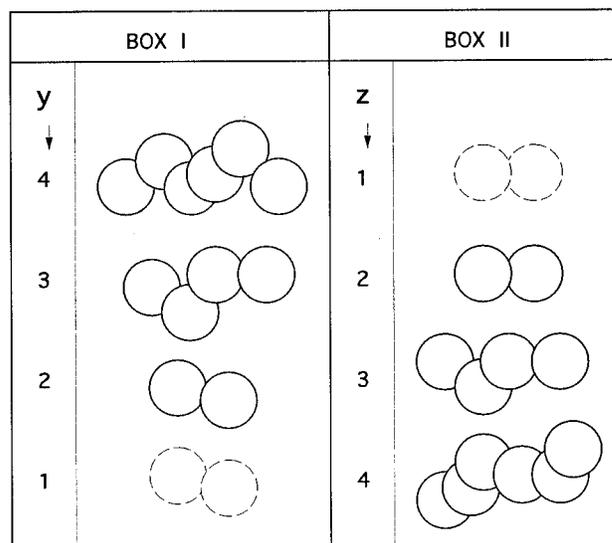


FIG. 1. Illustration of the correlation between states of the tagged chain in the two simulation boxes of an expanded Gibbs ensemble for a system of linear hexamer molecules. A choice of the intermediate states is shown here where transitions between neighboring states involve the simultaneous creation or destruction of two sites. To any given state on the left-hand side (box I) corresponds the state shown in the right-hand side (Box II). The dashed sites of state 1 represent a fully decoupled molecule. Note that, when assigning an initial state for a transition, states 1 and 4 are interchangeable.

incremental change in box I is accompanied by a decremental change in box II and viceversa. If  $y$  and  $z$  denote the current states in box I and II, then the following equality must hold:

$$y + z = M + 1. \quad (10)$$

Note that, as for the grand canonical case, states 1 and  $M$  are interchangeable as initial states (in each box); the nominal value is assigned depending on the type of transition proposed. The correlation of states ensures that the total number of sites in the system is preserved throughout the simulation. The probability of accepting correlated transitions in boxes I and II is given by the product of accepting the corresponding transitions independently. Since each box is representative of an expanded grand canonical (EGC) ensemble, it follows from Eq. (8) that for an incremental transition in box I we have

$$P_{\text{accept}} = \min \left[ 1, \frac{R_w^I}{R_w^{II}} \exp(\Psi^*) \right], \quad (11)$$

where

$$\Psi^* = \Psi_{y+1}^I - \Psi_y^I + \Psi_{z-1}^{II} - \Psi_z^{II}. \quad (12)$$

Substituting the  $\Psi$ 's from (6) into (12) we obtain

$$\Psi^* = \omega^* \beta \mu - \omega_{y+1}^I \log \frac{N_{y+1}^I}{V^I} + \omega_y^I \log \frac{N_y^I}{V^I} - \omega_{z-1}^{II} \log \frac{N_{z-1}^{II}}{V^{II}} \\ + \omega_z^{II} \log \frac{N_z^{II}}{V^{II}}, \quad (13)$$

where

$$\omega^* = \omega_{y+1}^I - \omega_y^I + \omega_{z-1}^{II} - \omega_z^{II} \quad (14)$$

and  $N_y^I$  is the number of molecules in box I including the tagged molecule if  $y > 1$ . In general, prior knowledge of the chemical potential of the system (i.e., a rough estimate) is useful for promoting a uniform sampling of the transient states; such information can be obtained from a short exploratory simulation. For simple homonuclear polymers, Eq. (7) provides a good prescription for the  $\omega$ 's. Combining Eqs. (7), (10), and (14) it follows that  $\omega^* = 0$  and, consequently, knowledge of the chemical potential can be forsaken. This approach proved to be satisfactory for the systems studied in this work. Note that if only two states are specified, i.e.,  $\omega_1 = 0, \omega_2 = 1$  ( $n_1 = 0, n_M = 1$ ),  $N_{y+1}^I = N^I + 1$ , and  $N_z^{II} = N^{II}$ , then the acceptance criteria Eq. (11) reduces to that for a CCB insertion-deletion attempt in a Gibbs ensemble. Further, if  $N_{sp}$  is also set to unity, then Eq. (11) becomes identical to the acceptance criteria for creation-destruction attempts in a conventional Gibbs ensemble.<sup>2</sup>

It is noted that in the EGC and EG ensembles, statistics for thermodynamic properties should in principle be collected only when the tagged chain visits the end states.

### III. RESULTS

For all the systems studied in this work we adopted uniformly distributed chain states such that  $n_{y+1} - n_y = 2$ . Conventional periodic boundary conditions were implemented around the simulation box. Molecular rearrangements were performed using single translations and extended continuum configurational-bias methods.<sup>17</sup> The number of trial positions per site for CCB incremental transitions varied from 6 at low densities to 200 at high densities.

A problem of great practical interest is the equilibrium partitioning of chain molecules between a pore and bulk fluid.<sup>18,19</sup> Unfortunately, grand canonical simulations of polymers in a bulk and in porous systems are extremely inefficient at liquidlike densities. Here, we present some results for hard-core flexible tetramers, hexadecamers, and 100-mers confined to a slitlike pore. EGC results for the bulk systems were completely consistent with previous data obtained by expanded canonical simulations (EVALENCH).<sup>9</sup> The pore walls are impenetrable and are  $5\sigma$  apart ( $\sigma$  is the bead diameter).<sup>18</sup> The length of the wall edges are  $12\sigma$ ,  $20\sigma$ , and  $100\sigma$  for the tetramer, hexadecamer, and 100-mer system, respectively. Our results are presented in Fig. 2 in the form of adsorption isotherms. The results of Yethiraj and Hall<sup>18</sup> for the tetramer system are also shown in Fig. 2; these are in good agreement with our data. These authors conducted canonical simulations in a large solvent bath in which a full pore was embedded. Such an approach is prone to significant finite-size errors and its applicability is limited to short chains. As previously noted, at low densities the chain prefers the bulk region to the pore; at high densities, the preference is reversed. The degree of partitioning is not affected significantly for the range of chain lengths studied in this work.

Expanded Gibbs ensemble simulations were performed for flexible homopolymer chains whose sites interact through

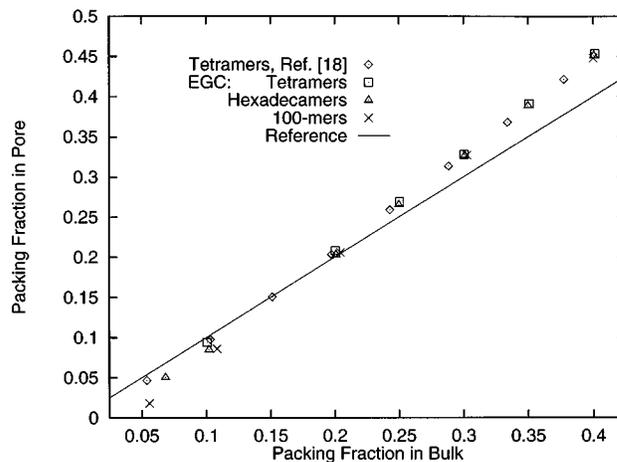


FIG. 2. Simulated adsorption isotherms for hard-sphere chains of 4, 16, and 100 segments in a slitlike pore. The pore walls are impenetrable and the gap between them is 5 site diameters. The diamonds depict simulation data for tetramers obtained by a canonical ensemble simulation (Ref. 18); the squares (tetramers), triangles (hexadecamers), and crosses (100-mers) show our results using expanded grand canonical simulations. The straight line corresponds to the case of equal densities in the two phases.

a square-well potential (well width =  $1.5\sigma$ ). The total number of molecules in the different systems was  $N_t = N^I + N^{II} \sim 150$  for  $n \leq 16$  and  $N_t = 80$  for  $n = 100$ . A simulation cycle comprised (in average)  $N_t$  molecular rearrangement attempts, a volume transfer move, and at least  $2 \times n$  transition attempts (evenly spread over the cycle). The total number of cycles varied from  $3 \times 10^5$  for tetramers to  $10^6$  for 100-mers. Figure 3 shows the results of our EG simulations. Coexistence data reported earlier<sup>20</sup> (using the more tedious iterative  $NPT-\mu$  method) are also shown in Fig. 3; good agreement is found between both sets of data. Using EG simulations we were able to obtain coexistence data closer to the critical point (although the uncertainty of the results is significant).

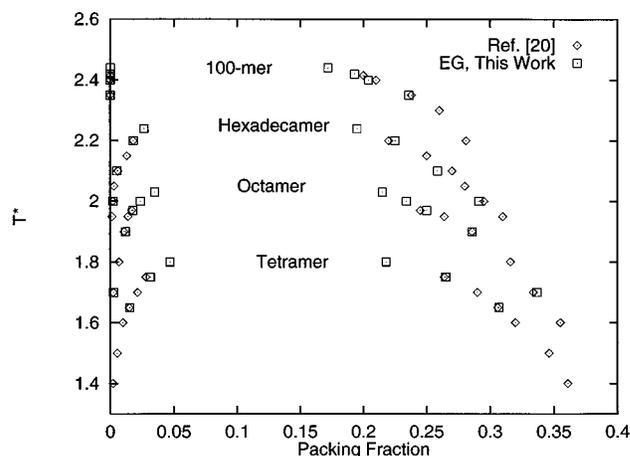


FIG. 3. Coexistence data for square-well chain molecules of 4, 8, 16, and 100 segments. The diamonds show simulation results using the  $NPT-\mu$  method as reported in Ref. 20 and the squares show results of expanded Gibbs ensemble simulations.

#### IV. PRACTICAL CONSIDERATIONS AND FUTURE WORK

Several parameters influence the efficiency of conventional grand canonical and Gibbs ensemble simulations (i.e., frequency and success rate of the different types of moves). Expanded ensemble simulations provide additional degrees of freedom which can be altered to aid simulation performance; preweighting factors, intermediate states, frequency of transitions, number of trial orientations  $N_{sp}$ , etc., can all be tuned. For the systems studied in this work, EG simulations were faster than conventional simulations in driving the system to equilibrium; no attempt was made, however, to optimize our sensible initial settings. A more extensive study of the comparative performance of our methodology will be presented in a future paper.

The most critical advantage provided by an expanded ensemble is that it permits simulation of very large molecules (long linear or branched polymers, proteins, etc.), which until now has presented an insurmountable obstacle for conventional GC and Gibbs methods. Such systems also pose some difficulties to an expanded ensemble approach because, as the number of sites in a molecule increases, the number of intermediate states also increases. This implies that only for a small fraction of a simulation run does the tagged molecule visit the end states (where the relevant statistics are collected). A simple solution to this problem is possible if we *relax* our physical picture of the system and allow intermediate states of the tagged chain to be *valid* representations of the system. In this way, statistics can be collected regardless of the transient state. Clearly, this artifact introduces negligible deviations in the thermodynamic limit of an infinite number of molecules in the system. For finite systems, such an approach may only be justified for certain (simple) molecular architectures and for certain properties. One such property is  $\rho^*$ , the number density of sites of a phase, which is the relevant property in GC and Gibbs ensemble simulations. Consider an EGC simulation of a system initially containing  $N$  full particles where a successful transition,  $y = 1 \rightarrow y = 2$ , has taken place (creation of sites of the tagged chain). For the subsequent set of events whose final event is the first hitting time of either  $y = 1$  or  $y = M$ , the number of sites in the system at any particular state is  $(nN + n_y)$ . For the analogous process whose first event was  $y = M \rightarrow y = M - 1$  (destruction of sites of the tagged chain), the (nominal) instantaneous site count is  $[nN - (n - n_y)]$ . Since both processes are equally likely, the average of sites of both processes (equally weighted) can be found from

$$V\rho_{\text{all states}}^* = \eta \langle N \rangle_{\text{end states}} + \frac{1}{2} [2 \langle n_y \rangle - n]_{\text{interm. states}} \quad (15)$$

Note that the end states have been accounted for in Eq. (15) by appealing to their indistinguishability. It follows from (15) that, if  $\langle n_y \rangle = n/2$ , ( $2 < y < M - 1$ ), then

$$\rho_{\text{all states}}^* = \rho_{\text{end states}}^* \quad (16)$$

Our selection criteria [Eq. (7)] satisfies the condition  $\langle n_y \rangle = n/2$ , provided that (i)  $n_{y+1} - n_y = \text{constant}$ , and (ii) each state is visited with the same frequency. Our methods and simulation results are consistent with such conditions (within practical margins); for the finite systems studied in this work, we find Eq. (16) to be remarkably accurate. This argument is trivially extended to the EG formulation. Although a correction term may generally be needed in Eq. (16), intermediate states and preweighting factors can always be chosen so that such a term becomes unimportant compared to the overall uncertainty of the results. More extensive simulations are under way (longer polymers, small systems) to demonstrate the usefulness of this procedure.

Extensions of the methods presented in this work can be easily implemented. For polymeric systems at high densities, it is advantageous to resort to a semi-analytical mapping of excluded volume<sup>21</sup> and gradual single particle insertions.<sup>7</sup> Expanded GC, semigrand canonical, and Gibbs ensemble simulation for mixtures can be implemented by combining the methods given in this work and those given in Ref. 3 in a straightforward manner.

#### ACKNOWLEDGMENT

This work was supported by an NSF-NYI grant to J.J.dP.

- <sup>1</sup>B. Widom, *J. Chem. Phys.* **39**, 2808 (1963); M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
- <sup>2</sup>A. Z. Panagiotopoulos, *Mol. Phys.* **61**, 813 (1987).
- <sup>3</sup>A. Z. Panagiotopoulos *et al.*, *Mol. Phys.* **63**, 527 (1988).
- <sup>4</sup>D. Moller and J. Fischer, *Mol. Phys.* **69**, 463 (1990).
- <sup>5</sup>K. Ding and J. P. Valleau, *J. Chem. Phys.* **98**, 3306 (1993).
- <sup>6</sup>P. Grassberger and R. Hegger, *J. Phys.* **7**, 3089 (1995).
- <sup>7</sup>R. D. Kaminsky, *J. Chem. Phys.* **101**, 4986 (1994).
- <sup>8</sup>Wilding and M. Müller, *J. Chem. Phys.* **101**, 4324 (1994).
- <sup>9</sup>F. A. Escobedo and J. J. de Pablo, *J. Chem. Phys.* **103**, 2703 (1995).
- <sup>10</sup>M. Barosova *et al.*, *Mol. Phys.* **87**, 423 (1996).
- <sup>11</sup>M. Mehta and D. Kofke, *Mol. Phys.* **86**, 139 (1995).
- <sup>12</sup>A. P. Lyubartsev *et al.*, *J. Chem. Phys.* **96**, 1776 (1992).
- <sup>13</sup>P. Attard, *J. Chem. Phys.* **98**, 2225 (1993).
- <sup>14</sup>S. K. Kumar, I. Szleifer, and A. Z. Panagiotopoulos, *Phys. Rev. Lett.* **66**, 2935 (1991).
- <sup>15</sup>F. A. Escobedo and J. J. de Pablo, *J. Chem. Phys.* **103**, 1946 (1995).
- <sup>16</sup>D. Frenkel, G. C. A. M. Mooij, and B. Smit, *Condensed Matter* **4**, 3053 (1992).
- <sup>17</sup>F. A. Escobedo and J. J. de Pablo, *J. Chem. Phys.* **102**, 2636 (1995).
- <sup>18</sup>A. Yethiraj and C. K. Hall, *Mol. Phys.* **73**, 503 (1991).
- <sup>19</sup>L. F. Vega, A. Z. Panagiotopoulos, and K. Gubbins, *Chem. Eng. Sci.* **49**, 1 (1994).
- <sup>20</sup>F. Escobedo and J. J. de Pablo, *Mol. Phys.* **87**, 347 (1996).
- <sup>21</sup>F. Escobedo and J. J. de Pablo, *Mol. Phys.* (in press).