

Monte Carlo methods

Contents

- *Basic principle*
- *Standard Monte Carlo Moves Involving a Single Simulation Box*
- *Moves Specific to multi-phase simulations (Gibbs Ensemble)*
- *Evaluation of the Chemical Potential by Test Insertions*
- *Statistical Bias Monte Carlo Moves*
- *Thermodynamic Integration for Fluid Phase Equilibria*
- *Thermodynamic Derivative Properties*
- *Henry Solubility Constant*

1 Basic principle

When applied to systems at thermodynamic equilibrium, Monte Carlo methods consider only positions, not velocities, as outlined by Ungerer [1]. The contribution of velocities to the partition function is determined analytically. Compared with Equilibrium molecular dynamics (EMD), Monte Carlo (MC) methods also allow for building statistical ensembles at thermodynamic equilibrium. As they do not follow the evolution of a system with time, they do not address the dynamic properties of matter, such as diffusion, viscosity, or thermal conductivity. In compensation, they may address important changes in the configuration space, such as the withdrawal or insertion of molecules in the system, which would be difficult to handle with molecular dynamics. They are, therefore, the privileged way to simulate sorption with the Grand Canonical ensemble or fluid phase equilibrium with the Gibbs ensemble.

This section does not cover Kinetic Monte Carlo methods used to simulate the dynamics of systems with identified transition states, nor do we consider chemical reaction equilibrium.

1.1 Partition Function in the Configuration Space

The canonical partition function for N identical particles may be expressed as

$$Q_{NVT} = \frac{1}{n^{3N} N!} \int_{p_i} \exp(-\beta K(p_i)) dp_i \int_{r_i} \exp(-\beta U(r_i)) dr_i \quad (1)$$

where K is the kinetic energy and U is the potential energy of the system. This factorization is possible because U depends only on positions r_i , and K depends only on momenta p_i , so the summations over p_i and r_i may be carried out independently.

The integral over momenta may be carried out analytically, and this results in:

$$Q_{NVT} = \frac{V^N}{\Lambda^{3N} N!} \int_{s_i} \exp(\beta U(s_i)) ds_i \quad (2)$$

[1] Philippe Ungerer, B Tavitian, and A. Boutin, Applications of Molecular Simulation in the Oil and Gas Industry: Monte Carlo Methods, (Editions Technip, 2005).

where the integration variables s_i are dimensionless positions and Λ is a temperature-dependent factor arising from the integration over momenta, called the de Broglie wavelength. The potential energy U is the sum of external and intramolecular potential energies. For monoatomic particles of mass m , the de Broglie wavelength is expressed as follows:

$$\Lambda = \frac{h}{(2\pi mk_B T)^{1/2}} \quad (3)$$

For polyatomic molecules, the de Broglie wavelength includes contributions from the rotation and internal degrees of freedom of the molecule, and N is the number of molecules. However, the knowledge of Λ is not needed for most Monte Carlo algorithms; we refer the reader to McQuarrie [2], where its expression can be found for simple molecules.

In the case of molecules involving internal constraints (such as imposed bond lengths) and flexibility (either bending or torsional), equation (1) still holds if the amplitude of bending and torsional movements is sufficiently small [3]. Eq. (2) is the basis of most Monte Carlo studies to date. In the configuration space (i.e., the space of positions), the probability density is:

$$\rho_{NVT} = \frac{1}{Q_{NVT}} \frac{V^N}{\Lambda^{3N} N!} \exp(\beta U(r_i)) \quad (4)$$

As a result of equations (1) and (3), it is sufficient to consider system *configurations* (characterized by positions only) instead of system *states* in phase space (positions and momenta) to build the statistical ensemble.

1.2 Markov Chain

The Monte Carlo technique uses a statistical method called a Markov chain for building the statistical ensemble. This method transforms the system from one configuration to another according to a given transition probability chosen to obtain the desired probability density. After a sufficient number of iterations, the system has visited a representative subset of the statistical ensemble, and the collection of visited system configurations may serve to compute average properties.

A Markov chain is characterized by the probability that a system in configuration a is transformed to configuration b . If this transition probability is noted as π_{ab} , the condition for the Markov chain to converge toward the probability density r is given by the following stationary condition, presented in matrix notation:

$$\rho\pi = \rho \quad (5)$$

where the dimension of the square matrix p and the vector ρ is the number of all possible configurations (a vast but finite number due to quantum mechanics principles).

In molecular simulation, we know the probability density of each configuration (for instance, the Boltzmann equation (4) in the canonical ensemble), and we must determine the elements of the transition matrix in such a way that equation (5) is respected. For this purpose, it is sufficient to use the following equation, known as the *microscopic reversibility* condition, for every pair of configurations a and b :

$$\rho_a \pi_{ab} = \rho_b \pi_{ba} \quad (6)$$

In this equation, the left-hand side is the flow of configurations a transformed into configurations b , while the right-hand side corresponds to the reverse flow from b to a . In practice, it is thus sufficient to define π_{ab} and π_{ba} in such a way that both flows are equal.

1.3 The Metropolis Algorithm

The Metropolis algorithm [4] is a way to generate a Markov chain where every step comprises two stages: in the first stage, a new configuration is generated by a random change (for instance, a random rotation

[2] D.A. McQuarrie, *Statistical Mechanics*, (Harper & Row, 1975).

[3] Nobuhiro Gō and Harold Scheraga, "On the Use of Classical Statistical Mechanics in the Treatment of Polymer Chain Conformation," *Macromolecules* 9, no. 4 (1976): 535-542.

[4] Nicholas Metropolis, Ariana W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller, "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics* 21 (June 1953).

or a random translation); in a second stage, the new configuration is accepted or rejected according to a criterion designed to generate the desired probability distribution. The probability that the new configuration is accepted is given by:

$$p_{acc}(old \rightarrow new) = \min\left(1, \frac{\rho_{new}}{\rho_{old}}\right) \quad (7)$$

where r stands for the probability density of the configuration in the statistical ensemble under consideration, and the acceptance probability p_{acc} is related to the transition probability of the Markov chain through:

$$p_{acc}(a \rightarrow b) = \frac{1}{\Omega} \pi_{ab} \rho_a \quad (8)$$

where Ω stands for the number of accessible configurations. The application of the Metropolis criterion (Eq. (7)) to the reverse flow (from the “new” to the “old” configuration) respects the microscopic reversibility condition stated by equation (6).

It is worth noticing that the Metropolis algorithm does not need to know the number of accessible configurations Ω to compute the acceptance probability. Also, the de Broglie wavelength Λ issued from the integration of the kinetic part of the energy often cancels out because only the ratio of the probability densities appears in Eq. (7). For instance, in the case of the *NVT* ensemble, the acceptance probability can be expressed as:

$$p_{acc}(old \rightarrow new) = \min(1, \exp(\beta(U_{new} - U_{old}))) \quad (9)$$

This criterion is used in the following way:

- $p_{acc} = 1$, i.e., $U_{new} \leq U_{old}$, the *new* configuration is accepted, i.e., it is added to the ensemble;
- if $p_{acc} < 1$, i.e., $U_{new} > U_{old}$, a random number q is generated between 0 and 1, and the *new* configuration is accepted if $q < p_{acc}$. Otherwise, the *old* configuration is added to the ensemble.

Monte Carlo methods require high-quality random number generators to apply the acceptance criterion.

Once the above procedure has generated a sufficient number of configurations, they form a representative subset of the statistical ensemble, i.e., every accepted configuration appears proportionally to its probability density. Then, standard averaging formulas can be used to derive macroscopic properties (potential energy, volume, pressure, etc.).

2 Standard Monte Carlo Moves Involving a Single Simulation Box

2.1 Translation

The first Monte Carlo move we consider is the *translation* of an individual molecule as a rigid body. It consists of the following steps:

- a translation vector $t = (t_x, t_y, t_z)$ is defined randomly, and a molecule is selected randomly,
- the translation vector t is applied to every atom of the molecule to obtain a new test configuration and the potential energy U_{new} of this new configuration is determined,
- the Metropolis acceptance criterion (9) is applied: if accepted, the new test configuration becomes the current configuration, and all variables (energy, etc.) are updated; if rejected, the old configuration remains the current configuration.

This move does not change the internal conformation of the molecule. In practice, the components of the translation vector are selected in a finite interval $-t_{max}, t_{max}$, which is smaller than the simulation box. In this way, the microscopic reversibility is respected, and it becomes possible to control the average acceptance rate of the translation moves.

Translations are sufficient to explore the entire configuration space in the straightforward case of monoatomic molecules in the *NVT* ensemble. As soon as more complex molecules or other ensembles are under consideration, additional Monte Carlo moves need to be introduced.

2.2 Rotation

The second type of move is the *rotation* of an individual molecule through a random angle α in a randomly chosen direction around its center of mass. The molecule is considered as a rigid body, i.e., bond lengths, bending angles, and torsion angles are preserved. As with translations, the random variation α is restricted to an interval $-\alpha_{max}, \alpha_{max}$ of controlled amplitude, and the criterion (9) is applied. Translations and rotations allow complete sampling of the *NVT* ensemble, provided the molecules under consideration are not subject to any internal deformations.

2.3 Volume change

In the *NPT* ensemble, a specific Monte Carlo move is used for *volume changes*, in which the simulation box is expanded or shrunk by an amount ΔV selected at random in an interval $-\Delta V_{max}, \Delta V_{max}$. In this move, the dimensionless positions of molecular centers of mass remain unchanged, as does the internal conformation of every molecule. The acceptance criterion incorporates the imposed pressure P according to:

$$p_{acc}(old \rightarrow new) = \min \left(1, \left(\frac{V + \Delta V}{V} \right)^N \exp(\beta(U_{new} - U_{old}) - \beta P \Delta V) \right) \quad (10)$$

where N is the total number of molecules in the simulation box.

2.4 Internal Moves

Internal Rotation

In order to sample internal changes in flexible chains, the simplest move is the *internal rotation* move (also known as *flip*), in which a randomly chosen atom of a chain is rotated around the axis formed by its two immediate neighbors, thereby preserving bond lengths and the BAC bond angle [5]. The amplitude is selected within a finite interval, and the acceptance probability is given by Eq. (9). This move is used to relax the inner part of a long linear chain molecule and flexible cyclic structures with United Atoms forcefields. Its limitation is that it cannot be applied as efficiently to branched molecules.

Reptation

Another very efficient MC move to simulate long-chain molecules is the *reptation* move. This move involves suppressing a segment of one or more atoms at one end of a randomly selected molecule and then adding an identical segment at the other end in a random position (taking care of possible constraints such as fixed bond lengths or fixed bond angles). Then, the acceptance criterion of eq. (9) is applied. The reptation applies only to linear molecules with a regular structure like ABBBBBBA or ABABABABAB, and the number of segments involved must be such that the inner part of the chain is unchanged (otherwise, computational efficiency and acceptance probability decrease significantly). Reptation is particularly efficient for long-chain molecules such as polymers [6].

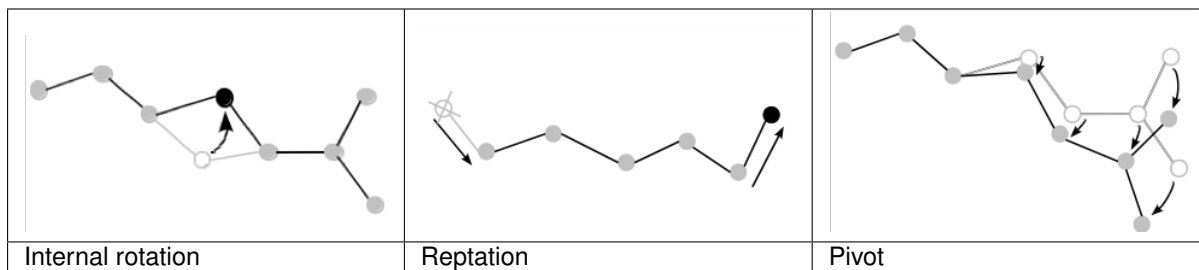
Pivot

When a molecule is composed of several more or less rigid parts (such as rings or branches) connected by a flexible chain, internal rotation and reptation moves only suffice to explore some possible internal configurations. Rotating part of the molecule around one of the atoms in a random rotation -the *pivot* move- makes more adequate sampling possible. The acceptance probability (9) is applied unchanged.

Monte Carlo moves available to sample internal conformation changes in linear or branched molecules (hydrogen atoms are not displayed)

[5] B R Dodd, T D Boone, and D N Theodorou, "A Concerted Rotation Algorithm for Atomistic Monte Carlo Simulation of Polymer Melts and Glasses," *Molecular Physics* 78, no. 4 (1993): 961-966.

[6] Benoit Leblanc, Bertrand Braunschweig, Hervé Toulhoat, and Evelyne Lutton, "Improving the Sampling Efficiency of Monte Carlo Molecular Simulations: an Evolutionary Approach," *Molecular Physics* 101, no. 22 (2003): 3293-3308.



Displacement, Partial Regrowth

In the *displacement* move (also referred to as translation and rotation in *MedeA GIBBS*), a molecule is deleted at its original place in the simulation box and inserted again at a randomly selected position in a randomly selected orientation and conformation.

The *partial regrowth* move applies specifically to flexible molecules, as it involves cutting one end off the molecule and allowing this part to regrow with a different shape.

These moves are generally used with statistical bias, which will be discussed later.

2.5 Insertion and Destruction Moves in the Grand Canonical Ensemble (GCMC Moves)

The fluctuation of the number of molecules is the characteristic feature of the Grand Canonical Monte Carlo simulation (GCMC). It is performed through two particular moves:

- the *insertion* of a new molecule. A new molecule of type i is tentatively inserted in a randomly selected location, with random orientation and internal conformation, unless a configurational bias is used. The insertion is accepted if the following Metropolis acceptance criterion is satisfied:

$$p_{acc}(insertion) = \min \left(1, \frac{V}{(N_i + 1)\Lambda_i^3} \exp(-\beta(U_{new} - U_{old}) + \beta\mu_i) \right) \quad (11)$$

where μ_i is the imposed chemical potential of molecular type i , N_i is the current number of molecules of type i in the simulation box, and the other symbols have the same meaning as in Eq. (2) and (9).

- the deletion of an existing molecule of the simulation box. After the selection of the molecular type i , a molecule is randomly chosen among those of type i in the current simulation box, and the following acceptance criterion is applied:

$$p_{acc}(insertion) = \min \left(1, \frac{N_i\Lambda_i^3}{V} \exp(-\beta(U_{new} - U_{old}) - \beta\mu_i) \right) \quad (12)$$

An equal number of insertion and deletion attempts are made for a given molecular type. Otherwise, the desired probability distribution is not satisfied. If we introduce the chemical potential $\mu_i' = \mu_i - \mu_{i_0}$ (where μ_{i_0} is the chemical potential of a perfect gas of pure compound i under a reference pressure P_0 at temperature T), the acceptance criteria can be expressed as:

$$p_{acc}(insertion) = \min \left(1, \frac{V}{(N_i + 1)\beta P_0} \exp(-\beta(U_{new} - U_{old}) - \beta\mu_i') \right) \quad (13)$$

$$p_{acc}(deletion) = \min \left(1, \frac{N_i\beta P_0}{V} \exp(-\beta(U_{new} - U_{old}) - \beta\mu_i') \right) \quad (14)$$

It is often assumed that the intramolecular energy is the same in the reservoir and the phase of interest, and then $U_{new} - U_{old}$ is the change in *external* potential energy.

In the section *Statistical Bias Monte Carlo Moves* moves, we will see how statistical bias may be used to increase the efficiency of these moves in dense phases, where the acceptance probability would be exceedingly low.

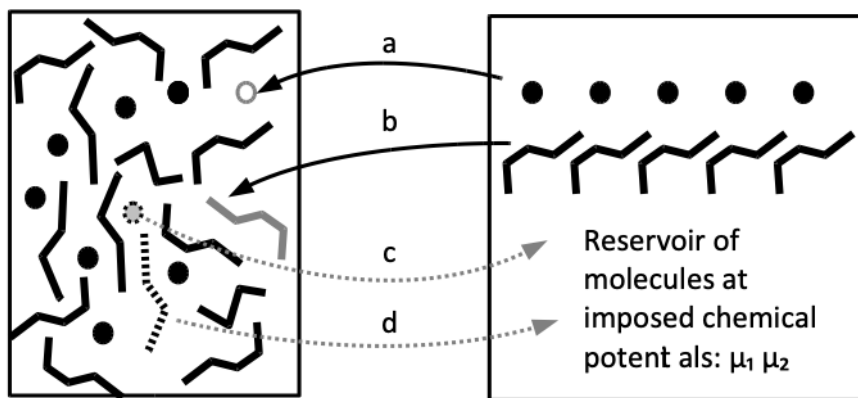


Figure 1: Insertion (a,b) and deletion moves (c, d) simulate the Grand Canonical ensemble of a binary mixture.

3 Moves Specific to multi-phase simulations (Gibbs Ensemble)

As the Gibbs Ensemble Monte Carlo (GEMC) involves more than one simulation box without them having an explicit interface, specific moves are used in addition to the single box moves discussed in the section *Standard Monte Carlo Moves Involving a Single Simulation Box*:

- transfers of molecules, which aim at imposing equal chemical potentials in all phases,
- coupled volume changes which provide mechanical equilibrium (equal pressures in all phases)

3.1 Transfers

The transfer move involves deleting a randomly chosen molecule in one phase *and* randomly inserting a molecule of the same type in the other phase in a joint move. The acceptance probability for the transfer of a molecule of type i from box A to box B is:

$$p_{acc}(\text{transfer } i_{A \Rightarrow B}) = \min \left(1, \frac{N_i^A}{V^A} \frac{V^B}{(N_i^B + 1)} \exp(\beta(U_{new}^A - U_{old}^A + U_{new}^B - U_{old}^B)) \right) \quad (15)$$

where the variables have the same meaning as in previous equations, superscripts A and B refer to the phase considered.

Similarly to insertion and deletion moves, GEMC transfers have a low acceptance probability when one or all phases are particularly dense or when large molecules are involved. In such cases, statistical bias methods are used to improve acceptance rates, as will be discussed later.

3.2 Volume Changes

In the Gibbs ensemble at imposed pressure, volume change moves are applied independently in every simulation box, and the acceptance probability is the same as shown in eq. (10). This ensures that the simulation is performed at the requested pressure.

In the Gibbs ensemble at imposed global volume $V = V^A + V^B$, phase A and in phase B volume changes δV in are applied. The acceptance probability is:

$$p_{acc}(\Delta V) = \min \left(1, \left(\frac{V^A + \Delta V}{V^A} \right)^{N^A} \left(\frac{V^B + \Delta V}{V^B} \right)^{N^B} \exp(\beta(U_{new}^A - U_{old}^A + U_{new}^B - U_{old}^B)) \right) \quad (16)$$

4 Evaluation of the Chemical Potential by Test Insertions

In contrast to volume, pressure, and energy, chemical potential cannot be obtained by simple averaging in the NVT or NPT ensembles but only by using test insertions in the system [7]. These tests are exactly like the GCMC insertions, except that they are never accepted, i.e., the list of molecules in the box is not updated. In the NVT ensemble, the chemical potential is thus obtained by the following relationship:

$$\beta\mu_i = \ln \left(\left\langle \frac{V}{(N_i + 1)\Lambda_i^3} \exp(\beta\Delta U^i) \right\rangle \right) \quad (17)$$

where ΔU^i is the variation of the total potential energy of the system when a test molecule of type i is inserted in it, and the brackets refer to an average over all test insertions in the NVT ensemble. When this method is applied to a dense fluid, most test insertions result in overlaps of the test molecule and existing atoms, so those high values of ΔU^i are found because of repulsion energy. The test insertions falling in free space between the fluid molecules correspond to negative values of ΔU^i , so they contribute more to the average.

The chemical potential μ_i appearing in equation (17) and the Boltzmann factor are not expressed in either the same units or with the same reference state as in classical thermodynamics. In these equations, μ_i is expressed in free energy per molecule and not per mole of substance, and both definitions differ thus by a constant $\ln(N_a)$, where N_a is the Avogadro number.

The classical thermodynamic reference state for chemical potential is a perfect gas of pure component i under a reference pressure P_0 and temperature T , which possesses a distribution of intramolecular energies but no intermolecular energy by applying eq. (17) to this reference state, it is possible to evaluate the chemical potential μ_{i_0} in the classical reference state. The classical chemical potential $\mu_i = \mu_i - \mu_{i_0}$ is then given by:

$$\beta\bar{\mu}_i = -\ln \left(\left\langle \beta P_0 \frac{V}{(N_i + 1)} \exp(-\beta\Delta U^i + \beta U_0^i) \right\rangle \right) \quad (18)$$

where U_0^i is the intramolecular potential energy in the reference state. (The difference $U^i - U_0^i$ is approximately the intermolecular potential energy if we consider that the intramolecular energy is similar in the reference state and the system of interest. The de Broglie wavelength Λ cancels out in this expression. The fugacity of component i can be readily obtained through:

$$f_i = P_0 \exp(\beta\bar{\mu}_i) \quad (19)$$

The Widom test converges very slowly in dense liquids or, more generally, in condensed phases. It is tempting to think of test deletions to compute chemical potentials, but this procedure would not appropriately sample the configuration space [8].

5 Statistical Bias Monte Carlo Moves

In order to increase the efficiency of sampling of transfer, insertion, deletion, and Widom test insertion moves, non-random MC moves are made, promoting sampling of favorable positions. Therefore, sampling is biased, and the expressions of acceptance probabilities must be corrected for the bias. Here, we will restrict the presentation to configurational bias and reservoir bias, which are presently available in *MedeA GIBBS*.

5.1 Configurational Bias

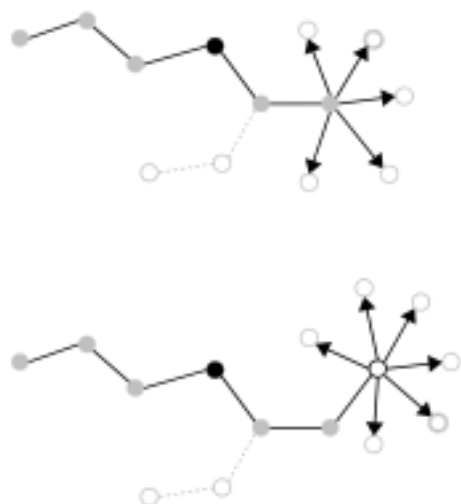
Configurational bias Monte Carlo (or CBMC) addresses the case of long linear or branched molecules that can adopt numerous conformations. This method takes advantage of the molecule's flexibility to grow it step by step, testing several possible random locations $k = 1 \dots k_{max}$ for the next atom, as illustrated below. The position of the new atom is selected according to the probability:

$$p(r_i) = \frac{\exp(-\beta u(r_i))}{\sum \exp(-\beta u(r_k))} \quad (20)$$

[7] B Widom, "Some Topics in the Theory of Fluids," *Journal of Chemical Physics* 39, no. 11 (1963): 2808.

[8] David A Kofke and P T Cummings, "Quantitative Comparison and Optimization of Methods for Evaluating the Chemical Potential by Molecular Simulation," *Molecular Physics* 92, no. 6 (1997): 973.

where $u(r_i)$ is the increment of potential energy associated with adding a new atom in position r_i . The exact process is applied to the next atoms until the end of the chain. Once the whole molecule has been regrown, the move is accepted according to a modified acceptance probability. In *MedeA GIBBS*, CBMC is used with insertions and deletions in Grand Canonical simulations, transfers in the Gibbs ensemble, Widom test insertions, partial regrowth, reptation, and displacements. The number of locations tested for each atom, k_{max} , can be different for all atoms of the molecule; e.g., a more significant number of test locations is often used for the first atom.



Configurational bias (CBMC) applied to the regrowth of the two last segments of a linear chain.

5.2 Reservoir Bias

For cyclic or branched molecules, the sampling efficiency can be increased by using a canonical reservoir of conformations of the molecule (or a part of the molecule), i.e., a collection of molecular conformations in which the Boltzmann distribution of internal energies is respected. In this way, the probability of occurrence of a given conformation in the reservoir is proportional to $\exp(-\beta U_{intra}^i)$.

In practice, this reservoir is created by performing repeated moves such as internal rotations, regrowth, etc.

In the case of Grand Canonical insertions, a molecular conformation is selected randomly in the reservoir and tentatively inserted [9]. The reservoir bias may also be used to improve the efficiency of CBMC algorithms with linear and branched molecules by picking bending angles from a suitable reservoir [10] instead of generating them repeatedly during the regrowth process.

In the case of Gibbs Ensemble transfers, reservoir bias can be exploited with an additional preliminary biasing step to find suitable “holes” in the liquid [11]. This involves the following stages:

1. in the first stage, several random locations for the center of mass are tested with a simple potential (i.e., a single Lennard Jones atom). One of these is selected with a probability as expressed in eq. (19), based on the interaction energy of the Lennard Jones force center with the system.
2. in the second stage, several molecular conformations are randomly picked in the reservoir and tentatively inserted in the system with the center of mass at the location identified in the first stage, in a random orientation. One of these is selected with a probability according to eq. (19), based on the interaction energy of the whole test molecule (or fragment) with the system.

[9] Jeffrey R Errington and Athanassios Z Panagiotopoulos, “New Intermolecular Potential Models for Benzene and Cyclohexane,” *Journal of Chemical Physics* 111, no. 21 (1999): 9731.

[10] Michael Macedonia and Edward Maginn, “A Biased Grand Canonical Monte Carlo Method for Simulating Adsorption Using All-Atom and Branched United Atom Models,” *Molecular Physics* 96, no. 9 (May 10, 1999): 1375-1390.

[11] Emeric Bourasseau, Philippe Ungerer, and Anne Boutin, “Prediction of Equilibrium Properties of Cyclic Alkanes by Monte Carlo Simulation New Anisotropic United Atoms Intermolecular Potential New Transfer Bias Method,” *Journal of Physical Chemistry B* 106, no. 21 (May 2002): 5483.

- the move is accepted or rejected with an acceptance criterion that corrects for the bias introduced in stages 1 and 2.

In the case of flexible cyclic molecules (such as cyclohexane), the reservoir bias move significantly saves computing time in condensed phases. It is also efficient for flexible branched molecules in which the possible conformations of branches are stored in a reservoir, avoiding the repeated sampling of high-energy bond angles.

In the case of rigid molecules, the reservoir of conformations is useless, but the two-step procedure outlined above significantly improves in GCMC and GEMC compared with the unbiased moves.

6 Thermodynamic Integration for Fluid Phase Equilibria

Thermodynamic integration is an efficient way to extend phase equilibrium calculations to lower temperatures. In the case of pure compounds, it is straightforward and efficient for obtaining vapor pressures at low temperatures, whereas Gibbs ensemble calculations would be no more tractable. This involves the Clapeyron equation:

$$\frac{\partial \ln(P_{sat})}{\partial \frac{1}{T}} = \frac{T \Delta_{vap} H}{P_{sat} \Delta_{vap} V} \quad (21)$$

where P_{sat} is vapor pressure, $\Delta_{vap} H$ is the change in molar enthalpy, and $\Delta_{vap} V$ is the change in molar volume accompanying vaporization in saturated conditions. The principle of thermodynamic integration involves estimating first P_{sat} at temperatures where Gibbs ensemble calculations are tractable and then using monophasic simulations of the liquid and vapor phases separately in the NPT ensemble to evaluate the right-hand side of eq. (21) at lower temperatures [12]. Thermodynamic integration may be simplified for temperatures significantly lower than the normal boiling point because the enthalpy and volume of a liquid depend little on the exact pressure value between 0 and 1 bar. Consequently, preliminary rough estimates of P_{sat} are sufficient to conduct the monophasic liquid NPT simulations.

7 Thermodynamic Derivative Properties

Thermodynamic derivative properties are second-order derivatives of the thermodynamic potential, as opposed to variables like pressure, molar volume, or enthalpy, which are first-order derivatives of the thermodynamic potential. Commonly used derivative properties include heat capacity, and isothermal compressibility, among others.

Derivative properties can be determined in principle from statistical fluctuations. As an example, it is straightforward to use equation (22) to determine the isothermal compressibility coefficient:

$$\beta_T = -\frac{1}{v} \left(\frac{\partial v}{\partial P} \right)_T \quad (22)$$

from Monte Carlo simulations.

Heat capacity cannot be obtained so simply because its calculation involves the fluctuations of the *total* energy E , while Monte Carlo results consider only the *potential* energy. Similar to equations of state, Monte Carlo simulations allow the determination of residual properties, i.e., the difference between the total and the ideal heat capacity. The molar residual heat capacity $C_P^{res} = C_P - C_P^{id}$ and the thermal expansivity

$$\alpha_P = \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_P \quad (23)$$

are derived from the following fluctuation formulae [13]

[12] David A Kofke and P T Cummings, "Quantitative Comparison and Optimization of Methods for Evaluating the Chemical Potential by Molecular Simulation," *Molecular Physics* 92, no. 6 (1997): 973.

[13] M Lagache, Philippe Ungerer, Anne Boutin, and Alain H Fuchs, "Prediction of Thermodynamic Derivative Properties of Fluids by Monte Carlo Simulation," *Physical Chemistry Chemical Physics* 3 (2001): 4333-4339.

$$C_P^{res} = \frac{\beta N_a}{NT} \left(\langle U^{ext} \bar{H} \rangle - \langle U^{ext} \rangle \langle \bar{H} \rangle + \frac{\beta N_a P}{NT} (\langle V \bar{H} \rangle - \langle V \rangle \langle \bar{H} \rangle) - R \right) \quad (24)$$

$$\alpha_P = \frac{\beta}{T \langle V \rangle} (\langle V \bar{H} \rangle - \langle V \rangle \langle \bar{H} \rangle) \quad (25)$$

where $H = U_{ext} + PV$ is the configurational enthalpy, U_{ext} the intermolecular potential energy, N_a the Avogadro number, and N the number of molecules.

The total heat capacity at constant pressure is obtained by adding the ideal heat capacity $C_P^{id}(T)$ to the residual heat capacity. The pure component ideal heat capacities, which are functions of temperature only, can be obtained from experimental correlations or group contribution methods [14] or QM calculations. For a mixture, C_P^{id} can be calculated from the pure components' ideal heat capacities through

$$C_P^{id}(T) = \sum_{i=0}^n x_i C_{P_i}^{id}(T) \quad (26)$$

where x_i is the molar fraction of each compound.

The Joule-Thomson coefficient, i.e., the derivative of temperature versus pressure at constant enthalpy, is expressed by the following equation [15]

$$\mu_{JT} = \frac{v}{C_P} (\alpha_P T - 1) \quad (27)$$

where v is the molar volume.

The molar heat capacity at constant volume, C_V , can be obtained from the Mayer equation:

$$C_P - C_V = \frac{\alpha_P^2 v T}{\beta_T} \quad (28)$$

The adiabatic (or isentropic) compressibility coefficient β_S is determined according to

$$\beta_S = -\frac{1}{V} \left(\frac{\partial \langle V \rangle}{\partial P} \right) = \frac{C_V}{C_P} \beta_T \quad (29)$$

The speed of sound in the fluid is obtained from the molar volume, the average molecular mass of the fluid, and the adiabatic compressibility coefficient

$$u_S = \sqrt{\frac{v}{M_w \beta_S}} \quad (30)$$

8 Henry Solubility Constant

The Henry solubility constant equals the solubility of a component in a solvent over its fugacity, in the limit of the infinite dilution:

$$K_{H_i} x_i = f_i \quad (31)$$

The Henry constant of light components in pure or mixed solvents may be obtained from the average chemical potential μ_i obtained by Widom test insertions through:

$$K_{H_i} = P_0 (N + 1) \exp(\beta \bar{\mu}_i) \quad (32)$$

where P_0 is the reference pressure for the chemical potential and N is the number of solvent molecules in the system. It is possible with a single MC simulation in the NPT ensemble to perform test insertions of several light components that are not present in the solvent. This allows the computation of several Henry constants in a given simulation.

[14] BE Poling, JM Prausnitz, and John P O'Connell, The Properties of Gases and Liquids, McGraw-Hill, 2001.

[15] R.A. Alberty and R.J. Silbey, Physical Chemistry, (Wiley, 1997).