2012

Vapor-Liquid Phase Equilibria Using Monte Carlo Wang-Landau Simulations

Tsvetan Aleksandrov

VAPOR-LIQUID PHASE EQUILIBRIA USING MONTE CARLO WANG-LANDAU SIMULATIONS

by

Tsvetan Dobromirov Aleksandrov
Master of Science, European Engineering School of Chemistry, Polymers and Materials Science, 2007

A Dissertation
Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota
December
2012

This dissertation, submitted by Tsvetan Aleksandrov in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done, and is hereby approved.

T. Demoskoski
Name of Chairperson

Name of Committee Member

Name of Committee Member

Name of Committee Member

Kathleen A. Hansen
Name of Committee Member

This dissertation is being submitted by the appointed advisory committee as having met all of the requirements of the Graduate School at the University of North Dakota and is hereby approved.

Wayne Swisher
Dean's name
Dean of the Graduate School

November 21, 2012
Enter the Date

PERMISSION

Title	Vapor-Liquid Phase Equilibria Using Monte Carlo Wang Landau Simulations
Department	Chemistry
Degree	Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Signature Tsvetan Aleksandrov

Date 12/10/2012

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am especially thankful to my research advisor, Dr. Jerome Delhommelle, for his ceaseless guidance, encouragement, and support throughout my studies at UND. During the course of this research, I benefited greatly through his knowledge and experience. He has a wealth of ideas and his enthusiasm is inspiring. He is exemplary, to students and colleagues alike, in his persistence, hard work, and adherence to high ethical standards. I am very appreciative of this opportunity to have had him as a mentor and example for an exceptional scientist.

I would like to thank my committee members, Dr. Kathryn Thomasson, Dr. Evguenii Kozliak, Dr. Lothar Stahl and Dr. Kanishka Marasinghe, for freely sharing their knowledge of chemistry with me, and offering me constructive criticism on my research and writing. I appreciate their invaluable time, guidance and teaching.

I am especially indebted to Dr. Caroline Deasgranges for her extensive support and guidance on my academic carrier. As a postdoctoral fellow in or research group she took time to advise and teach me and for that I am deeply appreciative. I would also like to thank all of my research group members, for sharing their knowledge with me.

I also extend my very special appreciation and gratitude to my devoted father and my mother and my loving relatives in Bulgaria and Sri Lanka for being with me in spirit, whenever I needed moral support. Lastly but certainly not the least, I want to thank my wife Duminda Liyanage Aleksandrova, for being a strong support in my life and

encouraging me throughout my PhD track.

ABSTRACT

The experimental determination of the critical point of metals is inherently connected with difficulties of measuring large temperatures and pressures. On the other hand, that of straight alkane chains larger than ten carbon atoms and branched chains larger than eight atoms is burdensome due to decomposition of the molecules while in the vapor phase. Therefore molecular simulations stand as an appropriate alternative for determining the critical point and phase coexistence properties of these fluids. Most simulation methods that are apt for this task rely on particles insertion and deletion moves that can significantly complicate the simulation, since high density in the liquid phases restrains this type of move. In our simulation this issue is avoided by combining the Wang-Landau (WL) sampling method in the NPT ensemble with efficient techniques for simulating dense liquid regions, such as configurational bias Monte Carlo (CBMC) and hybrid Monte Carlo (HMC). We simulate the liquid-vapor equilibria curves of copper and branched alkanes, isobutane and isopentane, by combining the HMC technique with the WL sampling. The standard boiling points of *n*-alkanes such as eicosane, tetracosane and triacontane are simulated by implementing the CBMC technique in the WL method in order to sample efficiently the various conformations of the long chain molecules. For copper we obtain a critical temperature T_c = 5695 ± 50 K, critical pressure P_c = 1141 ± 100 bar and critical density ρ_c = 1.80 ± 0.03 g/cm^3. All of these values lie within the range of experimental data. The vapor-liquid equilibria curves and critical

points of the branched alkanes are in excellent agreement with experimental data as well as simulation results using the Gibbs ensemble Monte Carlo method. The boiling points of the above listed *n*-alkanes also show good match with experimental data, with deviations in the range of 2 to 3 %. The WL simulations in the NPT ensemble is a simple and robust method for establishing vapor-liquid phase diagrams, as a single simulation run is necessary to evaluate the properties at a given temperature and for a large variety of substances. Its reliability is proven by the good agreement between experimental and simulation results.

CHAPTER I

INTRODUCTION

Depending on the arrangement of the building atoms or molecules, the materials can be found in four different types of phases - solid, liquid, gas and plasma. A phase is characterized with a uniform chemical composition and physical properties and well separated by other phases with boundary surfaces. The study of phase transformations is essential to control the desired characteristics of a given material. Thermodynamic parameters such as temperature and pressure are often used to drive these transformations in the forward or backwards directions and knowledge of the right amount of heat or work done to the system under study in order to control these parameters is required. It is convenient to present these parameters on a plot that indicates the phase domains, better known as a phase diagram (Figure 1.1).

At the boundary between two phases, a unique dependence between two or more of the thermodynamic parameters is established which controls the domain of the phase diagram where the two phases coexist. Under these conditions phase equilibrium is obtained and it is not underpinned by complete absence of change, but rather, the absence of the tendency towards change on a macroscopic scale. This tendency towards change is caused by a driving force and hence all the forces in equilibrium systems are exactly balanced. The origin of the forces could be mechanical work, which is controlled by the pressure exercised on the system. Heat change from one phase to the other would also

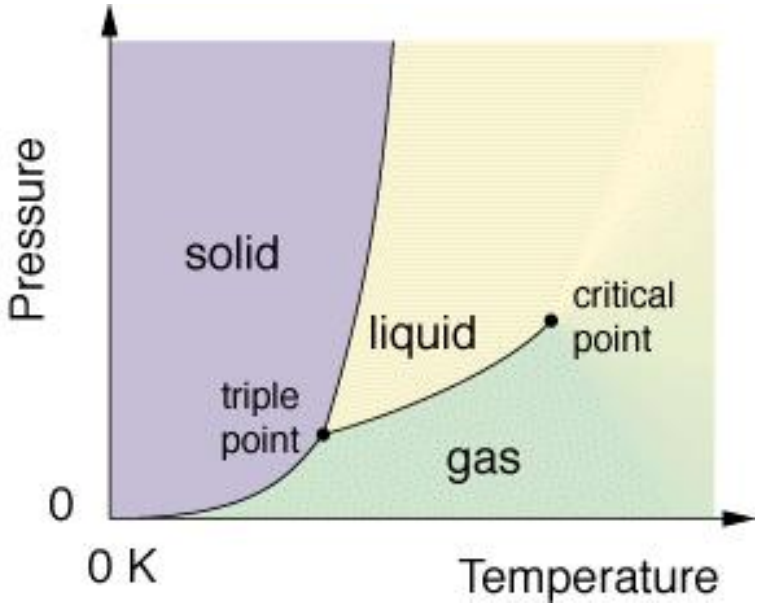


Figure 1.1. Example of a phase diagram

disrupt the equilibrium and that disturbance is caused by temperature difference. The imbalance of chemical potential on the other side causes particle transfer from one phase to the other. At equilibrium the thermodynamic variables that are equalized for the two phases are pressure, temperature and chemical potential.

The transition between the phases is known as first order phase transition. It is characterized with an abrupt change in the physical properties of the system. If we move along the line connecting the vapor and liquid phases, as temperature increases the difference between the densities of the two phases decreases. At the critical point it becomes zero and so it is possible to have a continuous transition from the vapor phase region to the liquid on the phase diagram (Figure 1.2). The differences in densities is non zero below the critical point and there is no continuous transition. This difference is called the order parameter of the vapor-liquid transition.

Phase equilibria of fluids is of significant importance in various fields of the biological and physical sciences. Fluids can be in several states of matter that are

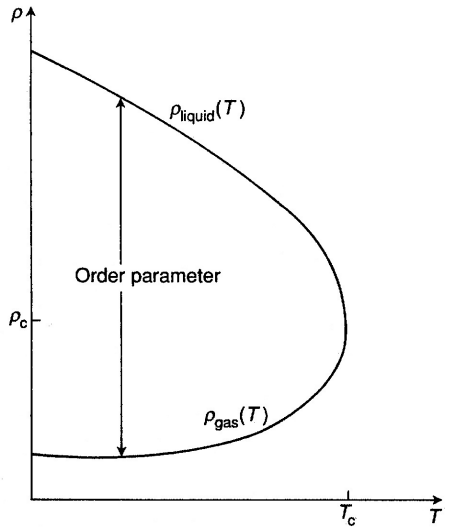


Figure 1.2. Example of a vapor-liquid equilibria curve . The difference in densities ($\rho_{vapor} - \rho_{liquid}$) is the order parameter.

subjected to continuous deformation under shear stress. As such, they exist mainly in the liquid and vapor phases. Transitions from one phase to the other is witnessed in a variety of physiological and industrial processes. In chemistry and chemical engineering, knowledge of the vapor-liquid phase envelope is crucial for many operations since often two or more phases are in contact. Examples include extraction, adsorption and distillation and the successful separation of mixtures lies in understanding the phenomena of phase equilibrium. A large number of experimental data has been produced to characterize the phase behavior of fluids. Furthermore, phase diagrams have been the subject of theoretical studies, mainly due to the necessity of equilibria parameters in equations of state (EOS). EOS relate state variables in an equation and as such are a powerful tool in predicting the properties of fluids. They are used extensively in engineering devices, materials and processes. Analytical solutions of the phase diagram, however, can only be obtained for simple fluids. A number of constraints is present in the experimental determination of phase diagrams when certain types of fluids are concerned.

1.1 Metals

Knowledge of the properties of metals in their solid phase is abundant, as most of them exist and are used for various purposes in this state. Little information is available, as to their characteristics in the liquid and vapor phase is concerned, due to the extreme conditions in which metals must be studied (Table 1.1). It is required in the nanosized top-down building approach, where vapor phase deposition techniques are used Other applications invlove geophysics and planetary science. Determination of the liquid-vapor phase diagram and critical properties, requires complicated experimental techniques, associated with a large investment in time and money, or estimation based on empirical or semi-empirical correlations.[26–28] Because of these issues, estimation of the critical properties can vary in a large diapason[29] and the critical points of only mercury and sodium have been measured experimentally.[30]

Table 1.1. Standard boiling points, critical temperatures and pressures of metals.[24]

Metal	T_{boil}, K	T_{crit}, K	P_{crit}, bars
Cu	2846	5140-5580	420-600
Ag	2437	3970-4480	150-270
Au	3081	7350-7610	2510
Al	2714	4000-4140	90
Zn	1184	3130-3240	1740
Cd	1040	2520-2570	1150-1230

1.2 Hydrocarbons

Certain hydrocarbons are also characterized with difficulties in the experimental evaluation of their vapor-liquid equilibrium properties and critical points. Alkanes are one of the main building blocks in biological molecules. They are also the main constituents in natural resources such as crude oil. Therefore their thermodynamic and transport

properties are of capital importance for the petrochemical industries. In large multicomponent mixtures, EOS facilitate significantly the prediction of their vapor-liquid equilibrium properties for optimization of the refinement process (Figure 1.3). A large number of EOS are parametrized based on the critical properties of the components and therefore knowledge of the critical temperature, pressure and density is required. For *n*-alkanes longer than ten carbon atoms or branched alkanes with more than eight carbon atoms however, the thermal instability of the chain renders the experimental estimation of the critical point an arduous task. Only a few techniques are capable of directly measuring the critical temperature and pressure of these alkanes, yet the integrity of the molecules after the experiment is sometimes questioned and no data of the critical density can be provided.[31,32] That is why molecular simulations stand out as a good alternative for studying such systems.

1.3 Molecular simulations of liquid-vapor phase equilibria

Molecular simulations have become an indispensable tool in the calculation of phase behavior of fluids. Simulation methods clearly do not suffer from the inconveniences of taking measurements at regions of instability of the molecules or under harsh condition at elevated temperatures and pressures. Furthermore computational data of the properties at equilibrium are increasingly used in the parameterization of transferable molecular force fields for more accurate prediction.[33–35] The computational determination of vapor-liquid phase equilibria relies on a statistical mechanics sampling of microscopic states along a path that connects the two phases in coexistence. The difficulty presented is in the efficient sampling of the interface region which is characterized with low statistical probabilities and high free energy. The task that is unfit

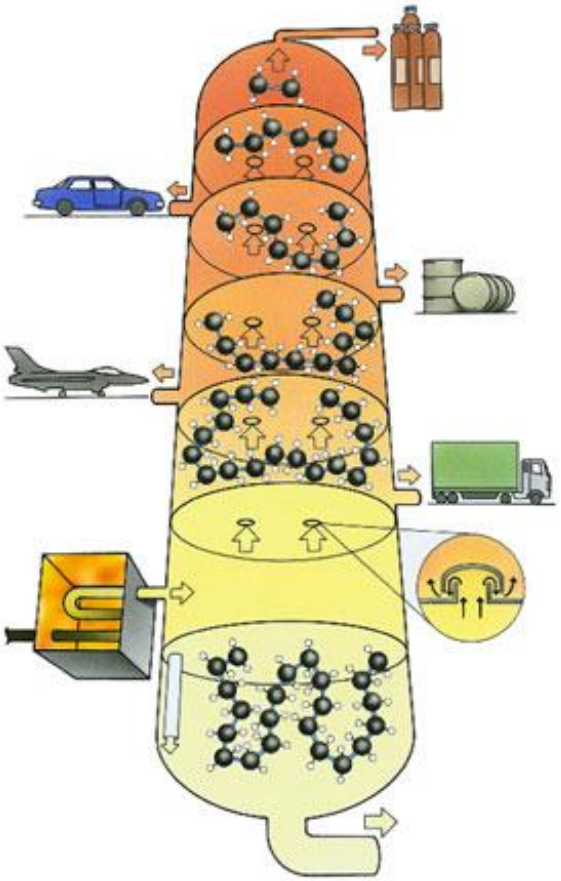


Figure 1.3. Schmeatic representation of crude oil refinement

for conventional simulation methods, based solely on Monte Carlo (MC) or molecular dynamics (MD) sampling, because exploration of the phase space is restricted to either side of the potential barrier, separating the two phases.

1.3.1 Gibbs ensemble Monte Carlo method

Advancements in the past two decades have allowed the calculation of properties at coexistence of the two phases. Gibbs Ensemble Monte Carlo (GEMC) method introduced by Panagiotopoulos, has gained a significant popularity in applications regarding vapor-liquid systems at equilibrium, due to its simplicity and robustness.[1] Unlike the simulations using MD and an equation of state, with this method simulations in

the liquid and vapor phases are carried out in parallel. This avoids the sampling of the interface region by performing the simulation in two distinct simulation boxes corresponding to the vapor and liquid phases (Figure 1.4). Even though separate, they are coupled through the MC algorithm, which ensures that both phases are in coexistence.

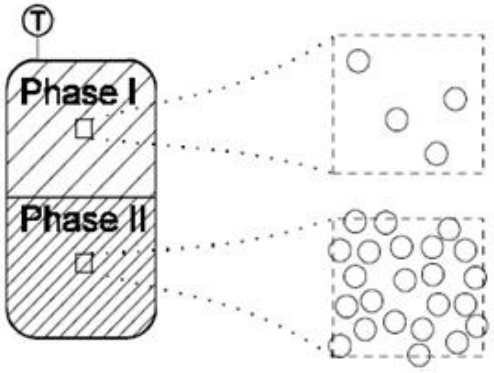


Figure 1.4. Simulation boxes representing the liquid and vapor state in the Gibbs ensemble [1]

The thermodynamic conditions of equilibrium for the vapor and liquid phases require that they have the same temperature, pressure and chemical potential. These requirements are met for the GEMC method by performing three types of moves in two simulation boxes, representing the two phases. First Monte Carlo (displacement) moves inside each of the box separately ensure that they are at internal thermal equilibrium. Random volume changes between the two boxes are carried out in order to ensure homogeneity of pressure. Finally, particle exchanges between the two boxes are performed to secure the same chemical potential, as part of the equilibrium criteria (Figure 1.5). Since the simulated system consists of these two boxes, the total number of particles and the total volume are fixed. The temperature is also fixed for the GEMC simulations as a part of the equilibrium requirement. The closest resemblance to the Gibbs ensemble is therefore the canonical ensemble with temperature, number of particles and

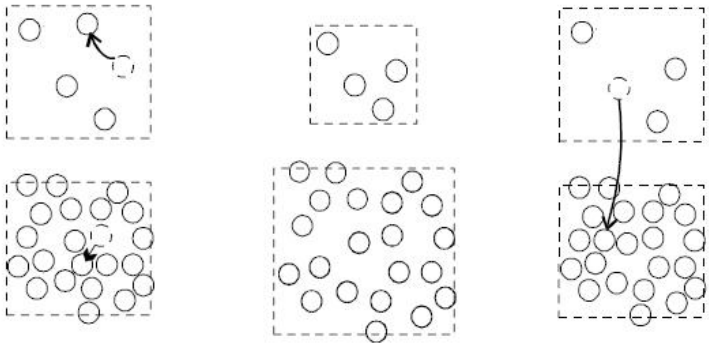


Figure 1.5. Type of moves characteristic of a GEMC simulation. From left to right the figures represent respectively Monte Carlo translational moves, random volume changes and particle exchanges [1]

volume fixed. The difference is that in the Gibbs ensemble the particle exchange moves and volume changes are performed in each of the subsystems (boxes), while keeping the total number of particles and volume constant. Thus if N and N_I correspond to respectively the total number of particles and number of particles in box I, then the number of particles in box II is $N_{II} = N - N_I$. The exact same dependence is valid for the volume ($V_{II} = V - V_I$). When the two phases coexist at temperatures and pressures away from the critical point, the equilibrium densities and compositions can be calculated by averaging the variables after equilibration.

In this example of equilibration, the two stable regions are characterized with a low density fluctuations for each phase. The free energy barrier between the two regions is the reason why the simulation remains in a stable state. If the two phases are closer to the critical point, this energy penalty for transition from one phase to the other decreases and frequent exchange of phases between the two boxes may occur. As a result the equilibrium densities cannot be determined accurately. Another issue related to the system approaching the critical point is that the correlation length of the system tends to infinity. The divergence cannot be simulated because of the finite sizes of the simulation boxes

thus once again forcing the two coexisting phases to be well below the critical point.

The GEMC method does not require prior knowledge of the chemical potential for the two phases at coexistence. As there is no interface separating the vapor and liquid phases, the high energy barrier does not present a constraint for direct calculation of the properties in both phases. This renders the method attractive for simulation of liquid and vapor phases at coexistence. The step involving the exchange of particles, however, could be burdensome in the case of high density liquids such as melted metals. The transfer from low density to high density states could often be rejected due to the highly unfavorable energetic states (particles overlap) in the liquid phase after the transfer. It is therefore necessary to carefully place the molecule from the vapor to the liquid phase, so that the particle insertion is accepted. In the case of chain molecules, configurational bias (CB) techniques allow the chain to be inserted bead-by-bead such that energetically favorable regions are found.[36,37] The combination of the two methods, known as Configurational Bias Gibbs Ensemble Monte Carlo (CB-GEMC), has been used in the construction of vapor-liquid phase diagrams of a variety of long-chain hydrocarbons.[38–41] Its main purpose is to enhance the acceptance rate for particle swaps between the two phases. Nevertheless the method can be applied only for chain molecules, while for molecules with more complex architecture other insertion techniques has to be considered.

The simulation of phase equilibria is very sensitive to the force field used to model the interaction between (intermolecular) and within (intramolecular) the molecules. In most cases the parameterization is achieved after fitting the parameters to the available experimental data. Comparison between the simulation and experimental results allows to assess the accuracy of the model as well as the parameterization. While some models are

capable of accurately predicting solely the liquid or vapor phase, when the two phases are simulated at equilibrium, they can significantly underestimate or overestimate the properties at coexistence (Figure 1.6).

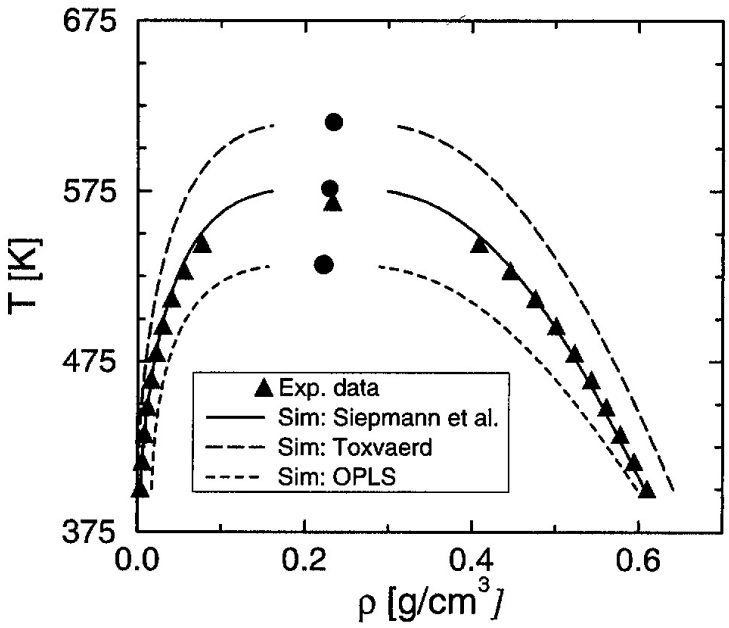


Figure 1.6. Vapor-liquid curve of octane using the GEMC simulations. Three different types of force fields are used - OPLS, model, the model of Toxvaerd and the TraPPE model of Siepmann[2]

In the example above (Figure 1.6), only the potential model developed by Siepmann is capable of accurately predicting the equilibrium densities and critical point of the molecule. This model was created on the base of previous models, such as these that underestimate and overestimate the experimental properties. The correct parameterization of the Siepmann model, made specially for fluids at coexistence has made it possible to obtain accurate results. The transferability of the model to molecules with common functionalities is achieved through comparison with experimental data and simulations of variable molecules with the same interaction potential sites. Therefore it is crucial that the simulation studies of variable molecules with the same potential are performed. The data

collected form the simulation are in terms used in algorithms that can modify the parameters of the potential so that the accuracy of the predicted values is improved.

1.3.2 Transition matrix Monte Carlo method

Another, more recent approach used in the estimation of vapor-liquid equilibrium curves is the transition-matrix Monte Carlo (TMMC) method.[42,43] In contrast to the GEMC method in which there is not a well defined ensemble, the algorithm developed by Errington can be used to simulate liquid-vapor phase coexistence in the grand canonical (GC) ensemble and a single simulation box is used.[15] When sampling occurs in this ensemble, the method is referred to as the grand canonical transition-matrix Monte Carlo (GC-TMMC). The constraints imposed on the system are the fixed chemical potential, volume and temperature (μVT) and two basic moves allow the system to be equilibrated: displacement of the molecule inside the system and particle exchange between the system and a thermal reservoir. As in the case of the GEMC method, displacement moves ensure the internal equilibrium inside the box. Particle insertion and deletion are performed to secure that the chemical potential is kept constant at the preliminary imposed value for the simulation (Figure 1.7).

The GC-TMMC method is based on the conventional MC simulations in the μVT ensemble and therefore it can be readily implemented by modifying the existing MC code for it. It utilizes the information about the attempted transitions between macrostates along a Markov chain. Similarly to the Metropolis scheme, the method also uses a detailed balance condition, but with macrostate probabilities. The macroscopic variable that changes during the simulation is the number of particles N, which ensures that both phases are visited during the course of the simulation. Therefore the statistical quantity

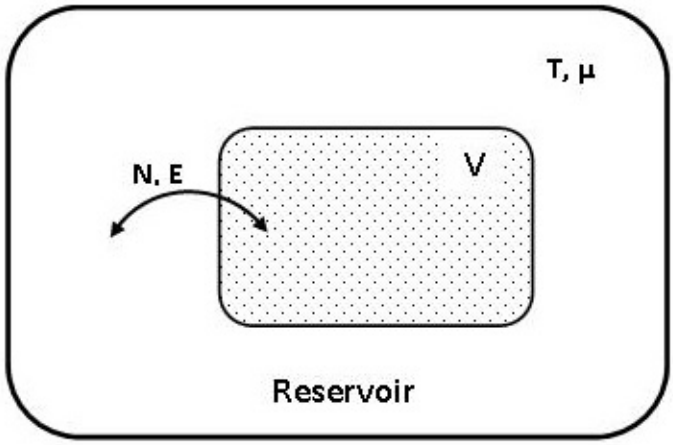


Figure 1.7. A simulation box used in the GC-TMMC method. Transfer of energy and particles is allowed, while the volume is kept constant as in the grand canonical (GC) ensemble.

that is traced is the probability of finding the system with a given number of particles $P(N)$, which in general can be obtained by summing over all the microstates with this particular number of particles. In the GC-TMMC scheme this is achieved by calculating an estimate of the macrostate transition probability of going from a macrostate with N number of particles to a neighboring one with either $N + 1$ or $N - 1$ particles ($\Pi(N \rightarrow N + \Delta N)$, where $\Delta N = \pm 1$).

The simulation starts by assigning a value to the chemical potential which is close to the value at coexistence and performing the two types of moves, while $P(N)$ is given an arbitrary value. In order to find $\Pi(N \rightarrow N + \Delta N)$, a collection matrix is created which contains the conventional MC acceptance probabilities of microstates associated with the two macrosates N and $N + \Delta N$ and updated after each move ($C(N \rightarrow N + \Delta N) = C(N \rightarrow N + \Delta N) + acc_{\alpha}$, with $acc_{\alpha} = \min[1, p_N/p_{N+\Delta N}]$). The collection matrix of comprising the acceptance probabilities of staying in the same

macrostate is also updated ($C(N \rightarrow N) = C(N \rightarrow N + N) + 1 - acc_{\alpha}$). The overall transition probability of moving to a macrostate with $N + 1$ number of particles, provided that the system is in a state with N ($\Pi(N \rightarrow N + 1)$) is calculated from the appropriately gathered collection matrix ($C(N \rightarrow N + 1)$) which is normalized. The probability of finding the system with the neighboring number of particles $P(N + 1)$ is then obtained from the detailed balance condition of the MC scheme ($P(N)\Pi(N \rightarrow N + 1) = P(N + 1)\Pi(N + 1 \rightarrow N)$). The microstates along the two phases are sampled by changing the number of particles and calculating the probability distribution at each step. This probability distribution is not normalized since an arbitrary value is set in the beginning of the simulation and the grand canonical partition function is not known.

Since sampling the two phases requires traversing a path of low probabilities, a weighing function is introduced in order to sample all states with equal probabilities. The conventional trial MC moves are accepted with a modified probability, which includes the macrostate probability ($acc_{\eta} = \min\left[1, \exp\left[\eta\,(N + 1)\right] p_{N+1} / \exp\left[\eta\,(N)\right] p_N\right]$ with $\eta(N) = -\ln P(N)$. This biased probability is used only to sample evenly the microstates, while the collection matrix is composed of the non-biased MC acceptance criteria. Initially the states that are sampled are consistent with the Boltzmann distribution for the values of μ, V and T specified. As the simulation progresses, $P(N)$ is periodically estimated using the collection matrices and hence more effective biasing function is obtained which helps to sample evenly all the states.

The probability distribution acquired with this method is subject to the chemical potential imposed. The value at which the two phases coexist (μ_{sat}) is found by shifting the originally taken chemical potential until the areas under the vapor and liquid probability

distribution domain are equal using histogram reweighting techniques.[44] At this point the areas under the vapor and liquid phases of the probability distribution have the same area.

The saturation pressure is determined by using the ideal gas as a reference state. This is necessary as no estimation of the partition function is provided. Similarly to the MC scheme, the partition function vanishes from the expression of the acceptance criterion as only the ratio of probabilities is calculated. As in the case of the GEMC method, without the implementation of a biasing strategy to insert the molecule in the liquid phase, the method is restricted to simple fluids with relatively low densities. The TMMC approach also relies on a pre-determined value of the chemical potential (or one close to it). That requires a series of short trial runs or extrapolation from known values at neighboring temperatures, which in certain cases can be a time consuming process.

1.3.3 Wang-Landau method

The solution to some of the issues presented in GEMC and TMMC techniques can be obtained by using the Wang-Landau (WL) sampling scheme. It was developed a decade ago by Wang and Landau as a tool to study first and second order phase transitions of two dimensional Potts and Ising model.[4] It allows for an efficient sampling of low probability states of vapor-liquid phases at coexistence that otherwise are infrequently visited by the Boltzmann criterion (Figure 1.8). This is achieved through the implementation of Monte Carlo simulations, combined with a biased distribution. The goal is to achieve an accurate estimate of a biasing function, which is essential in the calculation of thermodynamic properties.

Similarly to the GEMC and GC-TMMC, the method is simple to use, as a single simulation run is necessary to give rise to the properties at equilibrium. Moreover, it

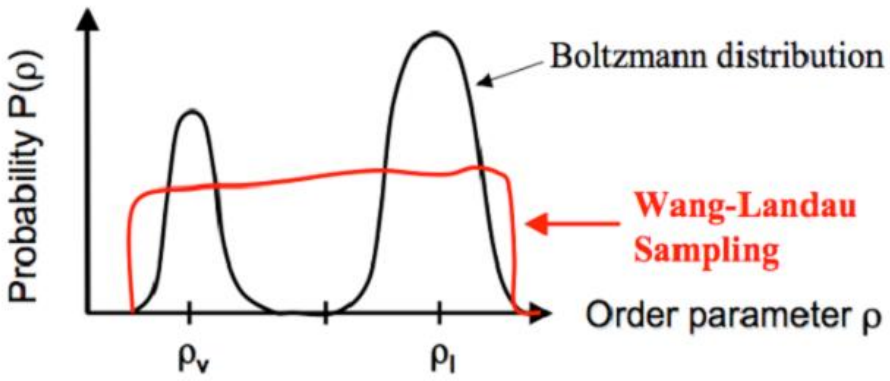


Figure 1.8. Boltzmann distribution (black line) and probability distribution obtained from Wang-Landau sampling (red curve)

allows the sampling of very high energy microstates at the interface through an accurate estimation of the partition function. The formalism in the isothermal isobaric (N, P, T) and grand canonical ensemble was developed by Ganzenmüller and Camp and used in the calculation of phase equilibria curves of monoatomic fluids.[45] The method has been successfully elaborated to simulate molecules with various architecture without the implementation of a an insertion bias, when simulations are performed in the N, P, T ensemble.[16,46]

The estimation of the biasing function stays at the heart of the WL method as a powerful tool towards the further evaluation of the Boltzmann probability distribution $p(X)$ and various thermodynamic quantities. While in the canonical ensemble the biasing function is the density of states, in the isothermal-isobaric and grand canonical ensemble its function is taken by the canonical partition function $Q_{N,V,T}$. It can be used to evaluate any thermodynamic property of interest such as density, entropy, Helmholtz free energy, Gibbs free energy etc, which gives the WL method a significant flexibility, when studying systems in equilibrium. These thermodynamic potentials cannot be directly calculated with the Gibbs ensemble and transition matrix MC methods, since they do not provide an

estimation of the partition function. It also allows to determine the interface region between the two phases, which is not accessible with the GEMC method. Furthermore when developed in the isothermal-isobaric ensemble, the WL method does not suffer from low acceptance rates of particle exchange typical for the two previous techniques and does not require an estimate of the chemical potential at equilibrium as in the GC-TMMC scheme.

1.4 Dissertation objectives and outline

The goal of this dissertation is to establish a vapor-liquid phase equilibria curves and estimate the critical properties of copper and alkanes by implementing the Wang-Landau simulation scheme. In chapter II, the historical background of the Wang-Landau sampling technique is presented with the various applications to which it can be extended in lattice systems. Furthermore, the description of the method is followed by its applications in off-lattice systems such as the calculation of vapor-liquid equilibrium curves of simple fluids as well as its extension to molecular fluids.

Chapter III presents the study of the vapor-liquid equilibria of copper. The formalism in the isobaric-isothermal ensemble is presented. A detailed description of hybrid Monte Carlo sampling is provided as well as its implementation in the WL scheme. The choice of the embedded atom model force field is discussed and compared to other existing force fields. During the simulation, the convergence of the partition function is traced as an important part of the accurate evaluations of properties at coexistence. The densities at coexistence are presented at temperatures close to the critical point in order to access the critical properties. The saturation pressures are evaluated at lower temperatures

and an estimate of the boiling point is presented. Results are compared to available experimental and simulation data.

In chapter IV, the problematics of the vapor-liquid equilibria studies for branch alkanes is presented. The force field model is presented and the implementation of the HMC scheme for the branched molecules is discussed. The difference with respect to the copper model are in the implementation of multiple time step algorithm for the efficient integration of bonding and non-bonding potentials. The results for two branched alkanes (isobutane and isopentane) are once again compared to experimental data as well as simulations using the GEMC method in order to establish the reliability of our simulation method.

Chapter V provides information on the Monte Carlo Wang-Landau scheme applied to long chain n-alkanes. The goal is to computationally determine the standard boiling points of these hydrocarbons, which are used in the parameterization of the transferable force fields. The conformational changes of the molecules are simulated using a configurational bias Monte Carlo scheme and its implementation in the WL method is explained in detail. The calculated data for the saturation pressure is presented and compared to experimental results to confirm the validity of the method.

Conclusions and suggested future work are summarized in chapter VI. The unique features of the WL scheme are presented with the promise for future applications.

CHAPTER II

WANG-LANDAU SAMPLING

2.1 Introduction

The Wang-Landau method is mostly similar to the multicanonical technique introduced by Berg and Neuhaus[47] for the study of first order phase transitions. In the latter method, an artificial sampling scheme is introduced in order to sample states that are seldomly visited with the conventional MC sampling. This is particularly important when visiting the low probability states at the interface between two phases at coexistence. The macroscopic states along the order parameter are sampled with equal probability, according to an inverse density of states. The density of states function is generated iteratively during several runs and is updated between runs, while at the same time the histogram of visited states is traced. The histogram gives information of the number of times a state is visited from where the density of states is estimated. The thermodynamic averages are obtained by averaging them with the Boltzmann probability distribution and the generated density of states. In the WL method the density of states is also the desired variable from the iteration scheme, however it is dynamically updated during each sweep allowing a more efficient even sampling of all the states.

2.2 Applications

2.2.1 Wang-Landau sampling of Ising and Potts models

In the original derivation of the method, Wang and Landau considered the 2D Ising and ten state Potts models.[3,4] The driving force behind the derivation of the method is the efficient sampling of phase space of systems with 1st and 2nd order phase transitions. The choice of the two mathematical models therefore is dictated by the fact that they serve as a starting point for new algorithms with applicability in a variety of problems in physics, chemistry and molecular biology, where short-range interactions between neighboring particles determine the behavior of the system. They present an excellent way of testing new methods. In these models the interaction sites are situated at the nodes of a two dimensional square lattice such as in Figure 2.1.

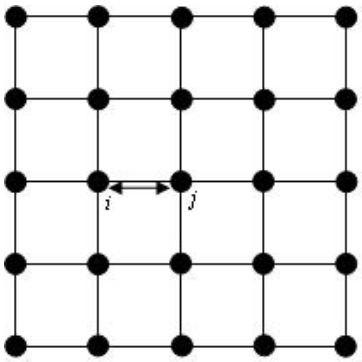


Figure 2.1. Schematic representation of the Ising and Potts models.

Only interactions between the nearest neighbors contribute to the total energy of the system. Thus only interactions between pairs of the nearest interaction sites are considered. To each interaction site in the Ising model an independent variable σ_i is assigned (i = 1.....N, where N is the total number of sites in the model). The variable σ_i can take only two possible values ±1, which build the energy space of the system (the

interaction sites are considered to have spins up (+1) and down (-1)). The Hamiltonian for these systems is written as:

$$\mathcal{H} = -\sum_{\langle i,j \rangle} J_{ij}\sigma_i\sigma_j \tag{2.1}$$

where J_{ij} is the unit energy of the neighboring interaction and equal to 1 for simplicity and the sum $\sum_{\langle i,j \rangle} J_{ij}$ is over all the neighboring pair interactions.

In the ten state Potts model, each site can have ten possible independent variables (q_i = 1....10). The Hamiltonian is given by the following expression:

$$\mathcal{H} = -\sum_{\langle i,j \rangle} J_{ij}\delta(q_i,\ q_j) \tag{2.2}$$

where $\delta(q_i,\ q_j)$ is the delta function and once again the unit energy J_{ij} is taken as one for each interaction.

Configurations in energy space are sampled by randomly selecting an interaction site on the lattice and flipping the spins for Ising model, while for the Potts model one of the ten possible variables is assigned to the site. If the sampling is performed without biasing the probability distribution, the histogram of the energy distribution will converge to the density of states in a very long run that visits all the possible spin states. For a 10x10 lattice, there are $2^{100} \approx 1.27\ 10^{30}$ states, while in the ten state Potts model there are 10^{100} states. The probability density for a given configuration in energy phase space is

$$p(\Gamma, E) = \frac{\exp(-E(\Gamma)/k_BT)}{Z} \text{ with } Z = \sum_{\Gamma} \exp(-E(\Gamma)/k_BT) \tag{2.3}$$

where Z is the system partition function and the sum is over all the possible configurations of energy phase space. The probability density at a given energy could be obtained by summing over the configuration with that particular energy

$$p(E) = \sum_{\Gamma \in E} \frac{\exp(-E(\Gamma)/k_B T)}{Z} = \frac{\Omega(E)\exp(-E/k_B T)}{Z} \tag{2.4}$$

where $\Omega(E)$ is the density of states or degeneracy of the energy level. If a biased probability density for a given configuration is created such that

$$p_{bias}(\Gamma, E) = \frac{p(\Gamma, E)}{p(E)} = \frac{1}{\Omega(E)} \tag{2.5}$$

the biased probability for a given energy level becomes the same for the whole energy phase space

$$p_{bias}(E) = \sum_{\Gamma \in E} p_{bias}(\Gamma, E) = 1 \tag{2.6}$$

Thus with a probability density, proportional to the inverse of the density of states (Equation 2.5), a flat histogram is generated for the energy distribution.

The simulation starts by assigning to all $\Omega(E)$ an arbitrary value ($\Omega(E)$=1 for all E). Upon visiting a given configuration, the probability at a given energy level is proportional to $1/\Omega(E)$ so that the acceptance probability for moving from an old to a new configuration becomes:

$$\text{acc}(o \rightarrow n) = min\left[1, \frac{\Omega(E_o)}{\Omega(E_n)}\right] \tag{2.7}$$

where the subscripts o and n denote, respectively, an old and a new configuration in

energy space. Each time a random change is induced, the corresponding new density of states is updated by multiplying the existing value with a modification factor $f > 1$ such that $\Omega(E_n) \rightarrow f\Omega(E_n)$. During the first sweep of obtaining a flat histogram, the f can be as big as the Napier's number e. If f is chosen to be too small, then a very long time will be necessary to reach all the energy states. On the other hand, a very large modification factor is also unfavorable since it will lead to large statistical errors.

After each sweep, the density of states converges to its true value with an accuracy proportional to lnf. The walk in energy space continues until all the states inside the allowed energy region are visited several times and the accumulated histogram of visited states $H(E)$ is relatively flat. A flat histogram as defined by Wang and Landau is such that the histogram $H(E)$ for all possible E is not less than 80% of the average one $\langle H(E) \rangle$. f is then reduced and the random walk in energy space is repeated until the same flatness criterion is met. This step is performed several times until the Boltzmann distributions between the sweeps do not differ significantly. This is achieved when f becomes smaller than some predefined value, close but greater than one. The manner in which f is reduced, thus has to be such that:

$$f_n \rightarrow 1 \quad \text{when} \quad n \rightarrow \infty \tag{2.8}$$

where n is the sweep number. The modification factor f is used as a control parameter for the accuracy of the density of states. The manner in which it is reduced should not allow it to become $f < 1$. In the original Wang Landau sampling it is decreased by the square root of its predecessor in each sweep ($f \rightarrow \sqrt{f}$). The closer it gets to one, the more accurate $\Omega(E)$. The modification factor is essential in controlling the convergence of the

simulation, since it determines the number of sweeps during the simulation and the accuracy of $\Omega(E)$. In the simulations of the two models, the final sweep of the random walk is such that $f_{final} = \exp(10^{-8})$. The final distribution of the density of states as obtained is only the relative density of states. The correct distribution can be calculated by either taking into account that the total number of possible states in the system or the number of ground states

$$\sum_E \Omega(E) = \begin{cases} 2^N & \text{for the 2D Ising model} \\ 10^N & \text{for the ten-state Potts model} \end{cases}$$

The number of ground states is 2 for the Ising model (all spins up or down) and 10 for the ten-state Potts model.

The detailed balance condition from which the acceptance criterion in Equation 2.7 is derived:

$$\frac{1}{\Omega(E_o)}\text{acc}(o \to n) = \frac{1}{\Omega(E_n)}\text{acc}(n \to o) \tag{2.9}$$

This condition is not exactly satisfied since the density of states is changed constantly during the random walk. Nevertheless, after several sweeps $\Omega(E)$ quickly converges to its true value as f approaches 1. Therefore it can be considered that the detailed balance condition is satisfied at the end of the simulation to within an accuracy proportional to lnf.

With the final distribution of the density of states, the WL method allows to access a broad range of properties. The canonical distribution at a given temperature can be found by simply multiplying the density of states with the Boltzmann factor:

$$p(E,T) = \frac{\Omega(E)\exp(-E/k_BT)}{Q} \tag{2.10}$$

Upon using the WL scheme, the canonical distribution can be found by the above formula. For these models the density of states is independent from temperature and thus a single simulation is enough to generate any temperature dependent distribution (Figure 2.2). The double peaked distribution in Figure 2.2 reflects that the transition temperature (T_c) is reached for the different sizes of the lattice system in the ten Potts model, since the peaks are characterized with equal heights. The conventional MC method cannot be applied to study the phase transition of such type as the depth of the valley between the two peaks serves as an impenetrable barrier for sampling states on the other side of the boundary. The depth increases with the size of the system, which can significantly limit the applicability of other methods for studying first order transitions.

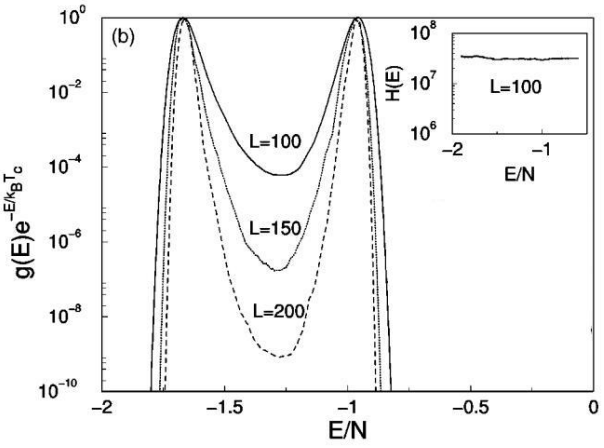


Figure 2.2. The canonical distribution at the transition temperature for different sizes (LxL) of the ten Potts model. The histogram of visited states for $L = 100$ is shown in the inset.[3]

The accuracy of the results from the simulation of the ten state Potts model is verified by comparing the transition temperatures for the different lattice sizes with

finite-size scaling theory:

$$T_c(L) = T_c(\infty) + c/L^d \tag{2.11}$$

where $T_c(L)$ and $T_c(\infty)$ are the first order transition temperatures for the finite and infinite size systems and d is the dimension of the lattice. The data from the simulation is plotted and is fitted according to the linear dependence of $T_c(l)$ as a function of L^{-d} as in Equation 2.11 (Figure 2.3).

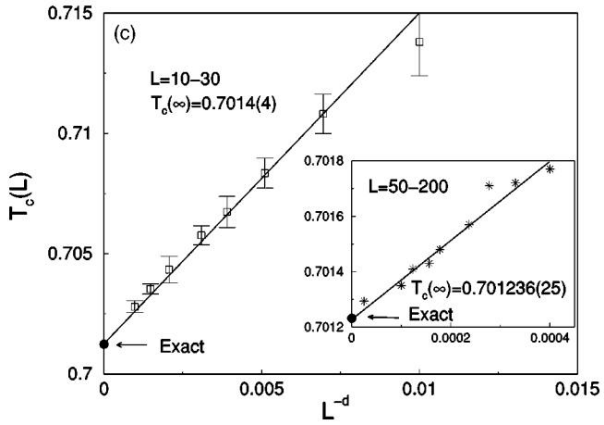


Figure 2.3. Transition temperatures from simulation data for the ten state Potts model with error bars obtained from multiple independent runs. Linear extrapolation for the $L = 10-30$ lattice sizes as well as for the $L = 50-200$ lattice sizes (inset).[3]

From the plot it can be noticed that the transition temperature for the infinite size system ($T_c(\infty) = 0.7014 \pm 0.00004$) is consistent with the exact solution ($T_{c,exact} = 0.0701232....$). In the inset an even more accurate estimate is obtained by enlarging the system size range from $L = 10 - 30$ to $L = 50 - 200$, with $T_c(\infty) = 0.701236 \pm 0.000025$, which is in excellent agreement with the exact solution. The WL allows one to calculate the density of states up to a 200×200 lattice model, while with the multicanonical approach, Berg *et al.* obtained data only for lattices up to

100×100.[47]

Another way of testing the accuracy and convergence of the WL method is obtained by comparing the simulation data to exact solutions for the 2D Ising model.[4] The density of states distribution from the simulation is presented in Figure 2.4 and compared to the exact results obtained by the Beale method.[48] It can be seen from the plot that no visible change can be detected between the two curves. Therefore in order to evaluate the accuracy of the WL method the relative error ($\varepsilon(X)$) is calculated

$$\varepsilon(X) = |(X_{sim} - X_{exact})\,/X_{exact}| \tag{2.12}$$

where X is a general quantity. Its distribution is presented in the inset of Figure 2.4 and it is as small as 0.035 % on the 32×32 lattice.

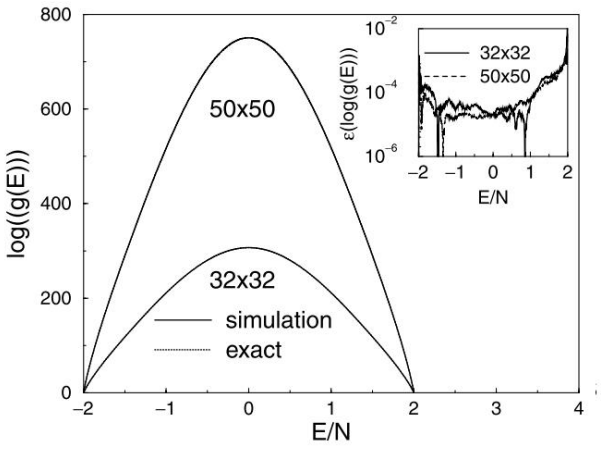


Figure 2.4. Density of states distribution obtained with the WL simulation and exact results obtained with Beale's method. The Inset shows the relative error ($\varepsilon(log[\Omega(E)])$).[4]

Similar results with low relative errors are obtained for the internal energy, free Helmholtz energy, specific heat capacity and entropy. These thermodynamic variables are calculated due to the evaluation of the the partition function:

$$Q = \sum_{E} \Omega(E) \exp(-E/k_B T) \tag{2.13}$$

The Helmholtz free energy is then given by:

$$F(T) = -k_B T \ \ln(Q) \tag{2.14}$$

The internal energy is

$$U(T) = \frac{\sum_E E\Omega(E)\exp(-E/k_B T)}{Q} \tag{2.15}$$

while the specific heat capacity can be calculated from the fluctuations of the internal energy

$$C(T) = \frac{\partial U(T)}{\partial T} = \frac{\left\langle E^2 \right\rangle - \langle E \rangle^2}{k_B T^2} \tag{2.16}$$

In the calculation of the canonical entropy, the Helmholtz free energy and the internal energy are used

$$S(T) = \frac{U(T) - F(T)}{T} \tag{2.17}$$

All of the above thermodynamic properties show almost a perfect match, compared to their exact values in a large range of temperature $k_B T = 0 - 8$ (Figure 2.5 and 2.6). As in the case of the density of states, there is no visible difference between the exact and simulated values and the relative error is shown in the inset. It rarely exceed 10^{-2}, which shows the reliability of the method when sampling lattice systems.

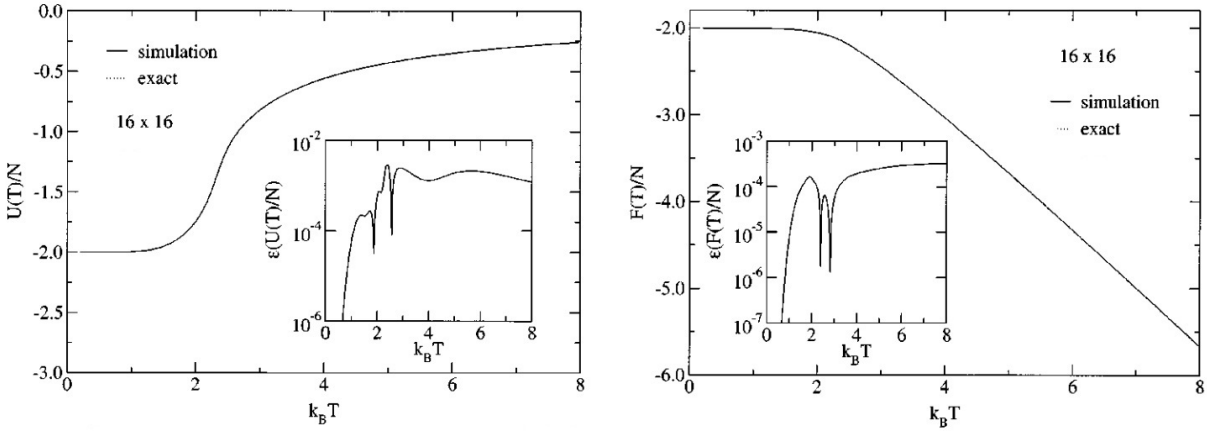


Figure 2.5. Internal energy (right) and Helmholtz free energy (left) as a function of temperature calculated by using the density of states.[5]

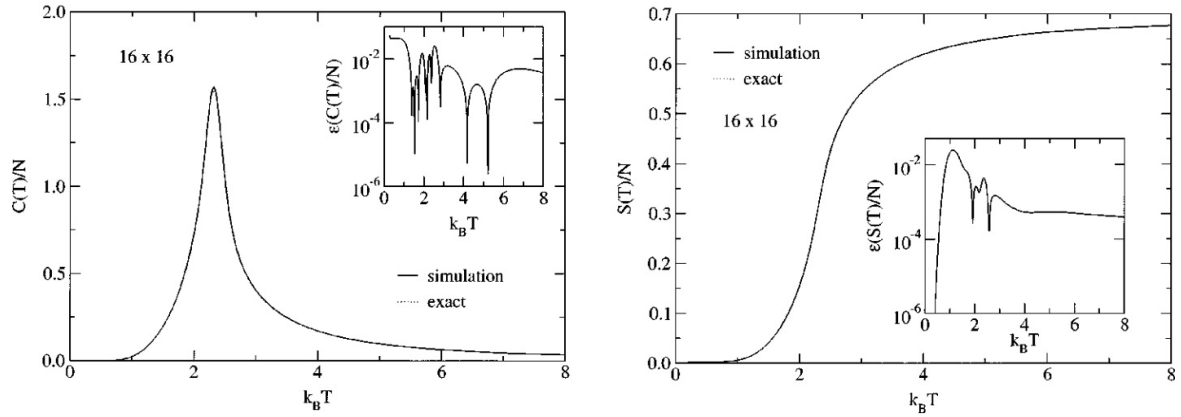


Figure 2.6. Specific heat capacity (right) and entropy (left) as a function of temperature calculated by using the density of states.[5]

In order to converge the results faster, Wang and Landau suggested dividing the energy region of interest into segments.[3] For each segment, the density of states is determined in the same manner as for the entire region, ensuring that a flat histogram is obtained for each. The overall $\Omega(E)$ is produced by joining together the segments. In the example given in Figure 2.7, the energy region $E/N \in [-2,\ 0.2]$ of 2D Ising model with $L = 256$ is divided into 15 segments and the density of states is estimated independently for each segment. Boundary effects are reduced by ensuring that each segment overlaps with its neighbors. The simulation can be speeded up this way, especially when performed

on multi-core processor.

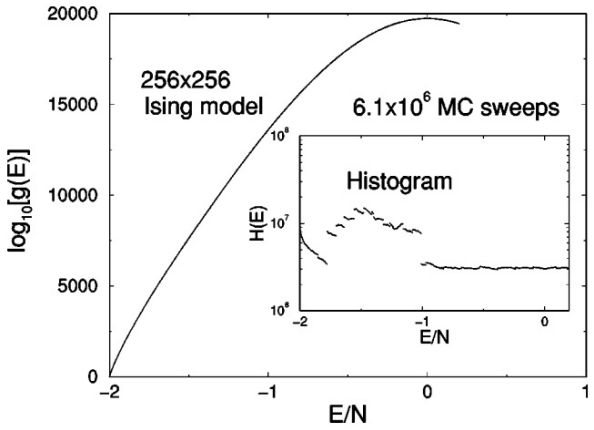


Figure 2.7. Density of states of the 2D ising model with $L = 256$ obtained after performing multiple walks in different segments. Inset - The flat histograms of overlapping segments.[3]

The problem related to this type of sampling is that larger errors of $\Omega(E)$ at the end of the sampled segment are observed. The sampling of energy states near the border of the segment can proceed in two different ways, in case the random flipping of the spins leads to an energy state which is outside the specified region. In Method 1 the attempted spin flip is rejected and the density of states is not modified with f nor is the histogram of visited states updated. In Method 2 the attempted spin flip is rejected and $\Omega(E)$ is modified with f ($\Omega(E) \rightarrow f\Omega(E)$) and the histogram is updated ($H(E) \rightarrow H(E) + 1$). Method 1 is used in Refs[3,4] which leads to systematic underestimation of $\Omega(E)$ at the borders of the energy values. The effect was later examined by Schultz *et al.* for 2D Ising model with $L = 32$ and for three different energy ranges. An overestimation of $\Omega(E)$ can be noticed at the right edge of the energy range (Figure 2.8). Such effect cannot be seen at the edges of each segment, but only at the one comprising the highest energy since a sufficient number of energy levels is allowed to overlap and the affected regions can be discarded when joining the segments. The overestimation at the right edge can be

explained by the non-updating of the density of states and histogram when a state outside the right edge is attempted. In the example given in the interval of interest $\Omega(E)$ increases monotonically with E. Because of this dependence, during the simulation, more states are sampled at the right edge of the energy range when using Method 1. The generation of states with energy higher than the right edge is more likely than the generations of states with energy lower than the left edge of the energy range. As a result, an overestimation of the density of states is observed only at the right edge. When $\Omega(E)$ is sampled according to Method 2, no such errors are present since both the density of states and the histogram of visited states are updated.

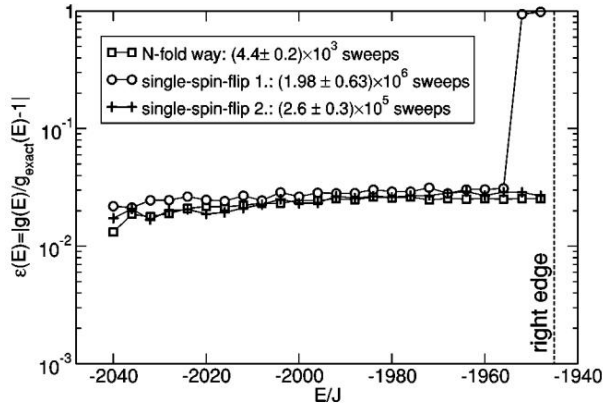


Figure 2.8. Relative error ($\varepsilon(E) = |(\Omega(E) - \Omega(E)_{exact})/\Omega(E)_{exact}|$) in the density of states for the first 25 energy levels in a 2D Ising model with $L = 32$.[6]

Another way of coping with the edge effects of the simulation for the Ising model was proposed by Schulz and coworkers.[7] The authors combine the WL algorithm with the N-fold way method by Bortz *et al.*[49] known for its better performance than the Metropolis algorithm at low temperatures. Because of the large value of the density of states, its decimal logarithm is considered $s(E) = \log_{10}\Omega(E)$. In the beginning of the simulation, $s(E)$ is set to zero for the energy energy range $E \in I = [E_{min}, E_{max}]$ where an estimation of

$\Omega(E)$ is desired. All the spins are divided into classes according to the energy difference ΔE_i that a flip of the spin can cause. For the 2D Ising model there are $M = 10$ classes ($i = 1.....10$). The overall probability P that a spin of class i is flipped is

$$P(\Delta E_i) = n(\sigma,\ \Delta E_i)p(E \rightarrow E + \Delta E_i) \tag{2.18}$$

where $n(\sigma,\ \Delta E_i)$ is the number of spins of state σ, belonging to class i. Similarly to the classical WL scheme he acceptance probability of visiting a new energy state $p(E \rightarrow E + \Delta E_i)$ is biased according to

$$p(E \rightarrow E + \Delta E_i) = \begin{cases} \min\left[1, \frac{\Omega(E)}{\Omega(E+\Delta E_i)}\right] & \text{if } E + \Delta E_i \in I, \\ 0 & \text{if } E + \Delta E_i \notin I. \end{cases}$$

The sum of probabilities for a spin flip within the for m classes is given by

$$Q_m = \sum_{i \leq m} P(\Delta E_i), \quad \text{with} \quad m = 1.....M \tag{2.19}$$

The value of Q_m is calculated in order to determine the class from which the flipping of the spin is done. A random number r is generated within the boundaries $0 < r < Q_M$ and a class m is selected if $Q_{m-1} < r < Q_m$. A spin from this class is chosen at random and as a result of the spin flip, the spins and its neighbors will change classes and the number of spins in state σ that belong to class i ($n(\sigma,\ \Delta E_i)$) will also change. In order to calculate the number of times the move made will be rejected on average in the thus proposed scheme (average life-timeτ), first the probability that the first random number will produce a flip is calculated by $\hat{P} = Q_M/N$. The probability that exactly n random numbers will produce a

new configuration is

$$\bar{P}_n = \hat{P}(1 - \hat{P})^{n-1} \tag{2.20}$$

The average life-time is then

$$\tau = \sum_{n=1}^{\infty} n\bar{P}_n = \sum_{n=1}^{\infty} n\bar{P}(1 - \hat{P})^{n-1} = \frac{N}{Q_M} \tag{2.21}$$

The steps that allow the estimation of the density of states are the following:

1. Initially the histogram of visited states as well as the weight factor $s(E)$ is set to zero ($H(E) = 0, \quad s(E) = 0$ for all E). The initial increment of Δs is chosen to be $\Delta s_0 = \log e$.

2. The acceptance probabilities $p(E \to E + \Delta E_i)$ are calculated and then the Q_m's of the configuration using Equations 2.18 and 2.19.

3. The average life-time is determined by using Equation 2.21.

4. The histogram of visited states as well as the density of states are updated according to

$$H(E) \to H(E) + \tau \tag{2.22}$$

$$s(E) \to s(E) + \Delta\bar{s} \tag{2.23}$$

$$\Delta s_i \to \Delta s_{i+1} \tag{2.24}$$

with

$$\Delta\bar{s} = \begin{cases} \Delta s_i \tau & \text{if } \Delta s_i \tau \leq \log_{10} e, \\ \log_{10} e & \text{if } \Delta s_i \tau > \log_{10} e \end{cases}$$

and

$$\Delta s_{i+1} = \begin{cases} \Delta s_i & \text{if } \Delta s_i \tau \leq \log_{10} e \\ \Delta\bar{s}/\tau & \text{if } \Delta s_i \tau > \log_{10} e \end{cases}$$

if Δs_i

5. The flatness of the histogram $H(E)$ is checked. It is considered flat if

$$H(E) \geq \varepsilon \langle H(E) \rangle \tag{2.25}$$

where ε is usually between 0.7 and 0.95.

6. Δs_j is refined according to $\Delta s_{j+1} = \Delta s_j / 2$.

7. Another spin is flipped as described and the steps are above are repeated starting from step 2.

One of the advantages of the N-fold way version of the WL scheme presented here is the low relative error $\varepsilon(E)$ of the density of states (Equation 2.12 with $X = \Omega(E)$). This is particularly noticeable at the right edge of the energy range in Figure 2.9. With the N-fold way, the systematic error at the right edge do not occur, contrary to the original

WL method (single spin flip). This shows that the N-fold way can overcome the issues related to the edge region of energy space where the acceptance rate for a spin flip is low. With this method, the density of states at the edges is sampled properly since in the definition of the acceptance probability ($p(E \rightarrow E + \Delta E_i)$), it is forbidden to sample states outside the predetermined energy range.

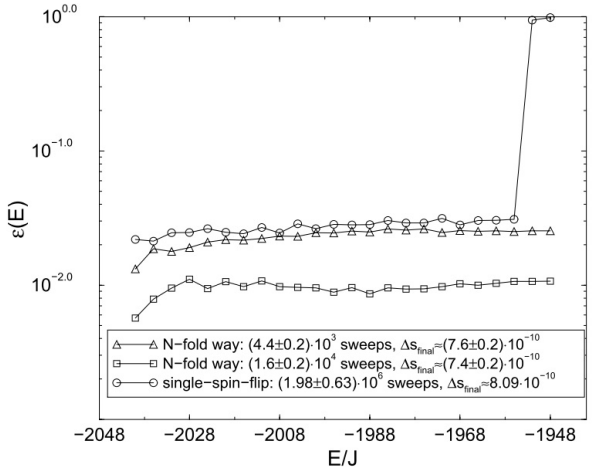


Figure 2.9. Relative error in the density of states $g(E)$ for the first 25 energy levels of a two-dimensional Ising model with $L = 32$.[7]

The method was applied also on thin Ising three dimensional films $L \times L \times D$. In this model the system is confined between two walls with fields h_1 and h_2 with periodic boundary conditions in the $L \times L$ planes (Figure 2.10).

The Hamiltonian for such a model is given by

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - H \sum_{i \in \text{bulk}} \sigma_i - h_1 \sum_{i \in \text{surface 1}} -h_2 \sum_{i \in \text{surface 2}} \tag{2.26}$$

where H is the internal field and as in the case of the 2D Ising model the spin can have a value of $\sigma_i = \pm 1$. For a system with dimensions 32×6, surface fields of $h_1/J = -h_2/J = 0.55$ and internal field $H = 0$, several thermodynamic properties are

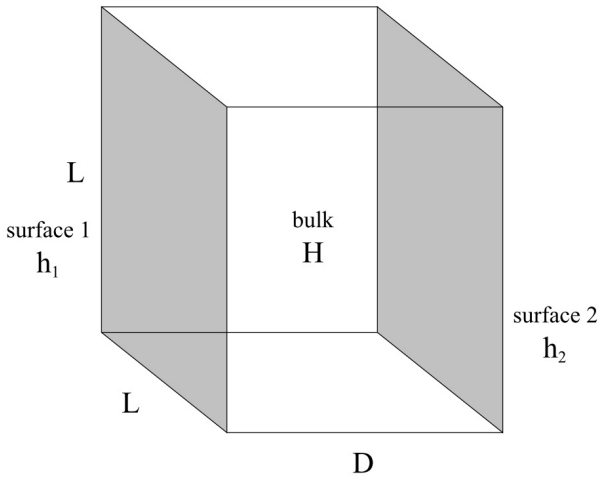


Figure 2.10. Thin film 3D Ising model with fields h_1 and h_2 acting on the surfaces and a field H on the bulk.[7]

calculated. In Figures 2.11 and 2.12 the plotted specific heat and internal energy as a function of the inverse temperature show a good consistency with simulation results from a standard single spin-flip heat bath algorithm.[50] In case of the specific heat capacity, the N-fold WL method even exceeds the results from the heat bath algorithm, the latter being characterized with more scattering around the guiding line.

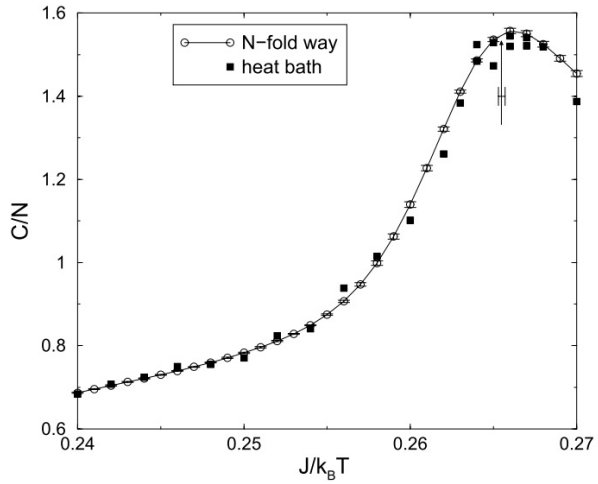


Figure 2.11. Specific heat capacity C/N versus the inverse temperature calculated with the N-fold WL method and a standard single spin flip heat bath algorithm.[7]

The evaluation of the density of states is particularly important since it allows for

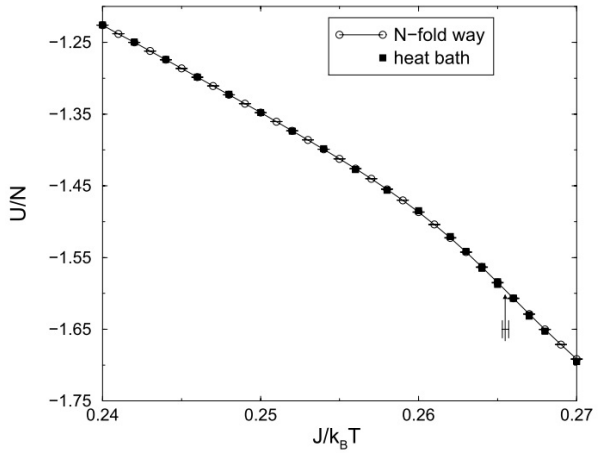


Figure 2.12. Internal Energy U/N versus the inverse temperature calculated with N-fold WL method and a standard single spin flip heat bath algorithm.[7]

the calculation of the partition function according to Equation 2.13 and consequently any thermodynamic property such as those shown above. This is an important advantage over the conventional MC simulations since for the latter the partition function cannot be accessed. The WL method as developed for the Ising and ten state Potts models is applicable to a wide range of systems. The method was further used to sample the phase space for other lattice models with applications to larger molecular systems.[8,10,51]

2.2.2 Other lattice systems

The WL scheme has been applied in the study of the energy barriers that separate glass and melt conditions in thin polymer films.[51] The lattice model is three dimensional with polymer chains movement represented by self-avoiding random walks. Similarly to the 2D Ising and ten state Potts models, the polymer chain interaction sites interact with their nearest neighbors with a fixed energy potential. The acceptance criterion with probability reciprocal to the density of states is slightly modified in order to take into

account the configurational changes of the polymer molecules. These changes require a bias move since at low temperatures or low densities conventional polymer trials fail to sample efficiently the energy space. Therefore the bias trials are incorporated in the WL scheme in the probability of proposing a transition ($\alpha(o \rightarrow n)$) of the detailed balance condition. This probability for the direct and reverse move is considered equal and therefore eliminated from Equation 2.9 in the original WL scheme. In the method proposed by Jain and de Pablo,[51] this equation is written as:

$$\frac{1}{\Omega(E_o)}\text{acc}(o \rightarrow n)\alpha(o \rightarrow n) = \frac{1}{\Omega(E_n)}\text{acc}(n \rightarrow o)\alpha(n \rightarrow o) \tag{2.27}$$

Thus the acceptance criterion becomes:

$$\text{acc}(o \rightarrow n) = min\left[1, \frac{\Omega(E_o)\ \alpha(n \rightarrow o)}{\Omega(E_n)\ \alpha(o \rightarrow n)}\right] \tag{2.28}$$

The configurational bias move involves choosing a random bead on the polymer, removing either side of the chain and regrowing it in a energetically favorable spaces. In the cubic lattice there are $k = 6$ possible orientations for a bead from the chain and one position j is selected from the available empty e_i trial positions with probability

$$p_i^j = \frac{1}{e_i} \tag{2.29}$$

The Rosenbluth factor R_F is then calculated by

$$R_F = \prod_{i=1}^{l} e_i \tag{2.30}$$

where l is the number of beads that are regrown and the subscript F signifies that this is the forward move. The Rosenbluth factor is calculated also for the reverse move R_R this time to retrace the old configuration. The acceptance criterion for the move is then

$$\text{acc}(o \rightarrow n) = min\left[1, \frac{\Omega(E_o)\, R_F}{\Omega(E_n)\, R_R}\right] \tag{2.31}$$

Because of the athermal nature of the WL simulation, a variety of properties can be calculated as a continuous function of temperature and thus extract more information from a single run. This renders the method much more attractive than the conventional MC simulations, where temperature dependence discreticizes the studied variable. Upon applying the conventional MC algorithm for the calculation of temperature dependent properties, the end configuration from the previous run has to be taken as a starting configuration in the next run in order to enable the system to relax fully instead of freezing it. As pointed out by Rathore *et al.* this issue is particularly prominent for temperatures near the glass transition.[8] Upon tracing the average energy as a function of the reduced temperature, the WL method provides a direct and continuous estimate of the energy distribution (Figure 2.13). The results are consistent with the annealing and parallel tempering simulation methods. By contrast, the last two can only provide the average energy at particular temperatures.

The same conclusion can be drawn from the comparison of entropy as a function of the reduced temperature (Figure 2.14).[51] With the WL method, a single run is enough to provide data at any temperature range while in the case of conventional canonical ensemble simulations the process is serial and 4 to 5 times longer.

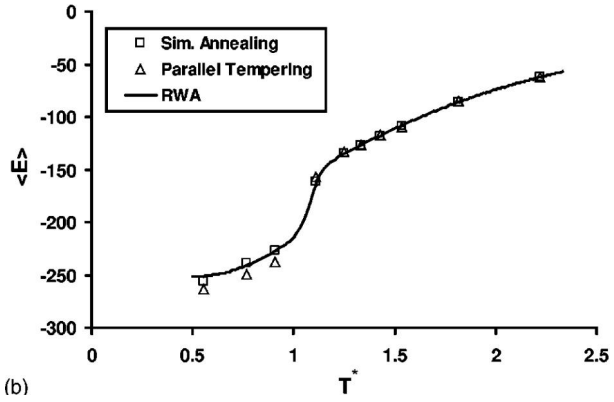


Figure 2.13. Average potential energy of a protein molecule as a function of the reduced temperature.[8]

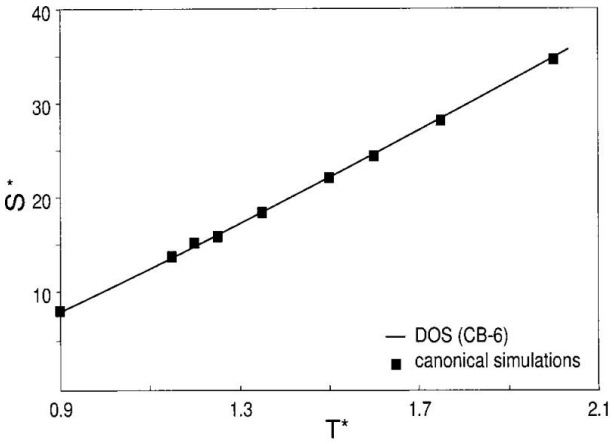


Figure 2.14. Average potential energy of a protein molecule as a function of the reduced temperature.[8]

Additionally, the WL scheme gives a direct estimate of the partition function in order to calculate entropy. This can be easily obtained through the value of free energy which, is a function of the density of states (Equation 2.14).

$$S^* = \frac{S}{k_B} = \frac{\langle E \rangle - F}{k_B T} \tag{2.32}$$

In the canonical ensemble, however, the evaluation of the free energy is calculated from the chemical potential, which on the other hand is obtained using the increment insertion method by Kumar and coworkers.[52] Thus WL simulations provide an important advantage

over conventional simulation methods in the complete thermodynamic characterization of the lattice system.

The WL scheme is a suitable scheme for studying folding processes of polymers and protein molecules. The potential energy landscape of these systems is irregular with high energy barriers and therefore only efficient sampling of low probability states such as WL sampling can be used in their simulations. It has been used to study the collapse of a single chain in the bond-fluctuation lattice model.[10] Monomers of the polymer chains are represented as unit cubes on a cubic lattice. The monomers are connected by bonds, which length can vary in a way that is compatible with the excluded volume between the monomers. The polymer chain is flexible and the shortest distance between two monomers is equal to 2 elementary cube sides while the longest one is $\sqrt{10}$ (Figure 2.15).

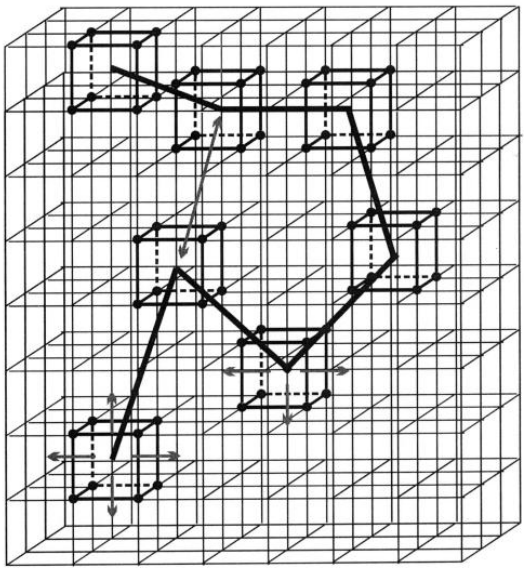


Figure 2.15. Schematic representation of the bond-fluctuation model on a cubic lattice. The double arrow stands for non-bonded interaction, while the single arrows represent the possible elementary motions of the monomers.[9]

A square well type potential is adopted acting between bonded and nonbonded neighbors,

$$E = -\varepsilon \frac{1}{2} \sum_{i} \sum_{j \neq i} \theta(6 - r_{ij}^2) \tag{2.33}$$

where $\varepsilon = 1$ is the unit energy and θ is the Heavistate function,

$$\theta(6 - r_{ij}^2) = \begin{cases} 0 & \text{if } r_{ij}^2 > 6 \\ 1 & \text{if } r_{ij}^2 \leq 6 \end{cases}$$

In the original WL scheme upon building the flat histogram of visited states, if an attempted new state is rejected, the old one is recounted. The density of states is again multiplied by the modification factor and, according to the acceptance criterion of Equation 2.7, this increases the probability of accepting the next new state with energy E_n. This propels the simulation to go through some hard-to-reach configurations in energy space. By doing so, it also creates an imbalance between the density of states for the old state ($\Omega(E_o)$) and that for all other energies ($\Omega(E^{'})$). In the coil-globule transition simulation of the bulk three dimensional polymer chain, the authors chose only to update $\Omega(E)$ only if the attempt is accepted, which led to the same density of states as the unmodified WL algorithm. The running estimate of $\Omega(E)$ is presented for each sweep of the simulation with decreasing modification factor f between each sweep (Figure 2.16). The bottom curve is the first iteration with the lowest value of f. As the simulation proceeds the shape of the curve becomes smoother and a more accurate dependence of $ln(\Omega(E))$ is obtained. After the first ten refinements of f it becomes indistinguishable on the scale of the plot, because the modification factor is close to 1. The absolute value of the density of states also increases with the number of sweeps and the real density of

states has to be corrected by a multiplicative factor, which obeys the condition that there is a single state when the energy is equal to zero ($\ln \Omega(0) = 0$).

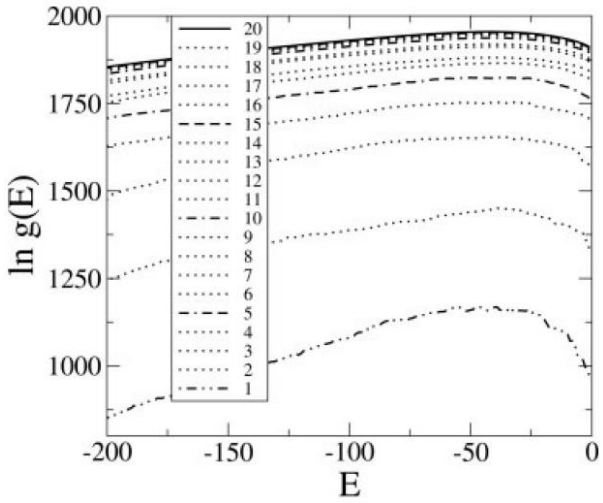
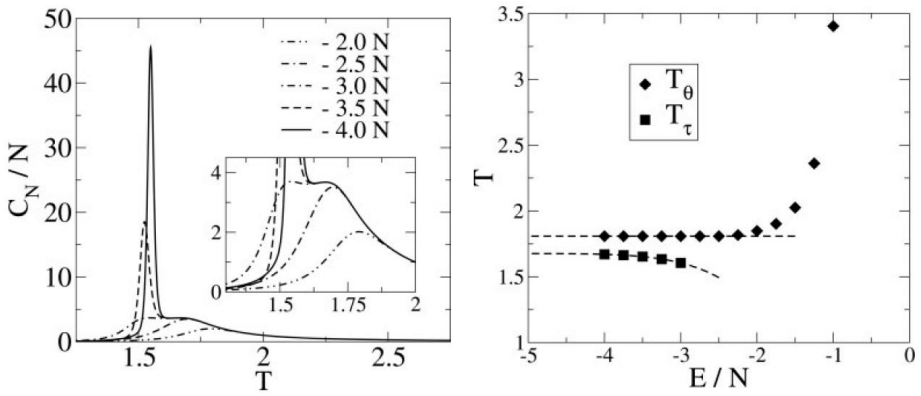


Figure 2.16. Running estimate of the density of states of a chain of length $N = 64$ with decreasing modification factor ($\ln f_1 = 1$ to $\ln f_{20} = 10^{-6}$).[10]

The determination of the second order coil-globule transition and the first order liquid-solid transition of the polymer chain is strongly dependent on the energy range considered. The shape of the curve tracing the specific heat (calculated using Equation 2.16) per monomer C/N as a function of temperature, changes for different ranges of the energy (Figure 2.17(a)). For a chain of length $N = 256$ monomers with $E_{min} = -2N$ and $E_{min} = -2.5N$ only one peak is visible. This range can be simulated using the conventional canonical MC sampling. However this is impossible for larger energy ranges, where E_{min} reaches $-4N$. This turns out to be an important issue as the peak indicating the second order coil-globule transition starts to shift its position to the left and a second peak appears indicating the first order liquid-solid transition. Thus to determine the correct temperature for both transitions, it is important to sample over the appropriate energy range. The stabilization of the peak position as a function of the energy range is presented in Figure

2.17(b) for $N = 512$ monomer units. The position of the peak corresponding to the coil-globule transition stabilizes around E_{min} is close to $-3N$. For the first order transition this happens at $E_{min} \approx -4N$. Thus, the accurate estimation of the transition temperature for the crystallization process (T_τ) and that of the coil-globule transitions (T_θ) strongly depend on the energy span chosen for the simulation.



(a) Specific heat as a function of temperature for different value of the minimum energy (E_{min})

(b) Shift in the transition temperature (T_τ - crystallization and T_θ - coil globule transition as function of the energy span

Figure 2.17. Transition temperature as a function of the sampled energy range[10]

2.2.3 Off-lattice systems

The WL method was originally designed to study systems with natural discretization of the energy phase space. The extension to the off-lattice systems is non trivial since the lattice systems are characterized with a constant composition and are assumed to be incompressible. Furthermore, the density of states is calculated to within a multiplicative constant. However, real continuum systems, such as fluids, are not incompressible and the absolute density of states is requires in order to calculate the thermodynamic properties. The generalization of the WL scheme to such systems was

proposed by Shell *et al.*[11] and applications involve mainly the construction of phase diagrams of vapor-liquid equilibria.[12,13,53]

In contrast to the lattice systems, the simulation of two phases in the real fluids cannot be performed by sampling solely the energy space. Since density is the order parameter, the energy sampling has to be accompanied with either volume changes or particles insertions and deletions. The density of states becomes dependent on either the volume or the number of particles together with the energy. The microcanonical partition function ($\Omega(N,\ V,\ E)$) is the appropriate density of states variable, because it gives the number of states with a given energy E, in a volume V for a number of particles N. Since the WL method is based on the MC random sampling of the configurational part of the phase space, it is convenient to work with the configurational part of microcanonical partition function ($\Omega_{config}(N,\ V,\ U)$ where U is the potential energy). According to the WL scheme, the probability of observing a particular configuration is inversely proportional to the density of states. In a continuum system this probability is:

$$p(N, V, U(q^{3N})) = \frac{1}{C}\frac{\mathrm{d}q^{3N}\mathrm{dV}}{\Omega(N, V, U(q^{3N}))} \tag{2.34}$$

where C is a normalization constant and q^{3N} is the generalized positions of the particles. The potential energy U is function only of the positions of the particles. The moves involved in the sampling of the phase space in real fluids are the particle displacement, volume scaling moves or particle insertion and deletion. The acceptance criteria for these moves are derived from the detailed balance condition. For a transition from an old (o) to a new (n) state the general acceptance criterion is

$$\text{acc}(o \rightarrow n) = min\left[1, \frac{p(n)\alpha(n \rightarrow o)}{p(o)\alpha(o \rightarrow n)}\right] \tag{2.35}$$

For a displacement move, a particle is taken and placed randomly within a cubic volume with side twice the maximum displacement distance in each direction (δ_{max}).

$$\begin{aligned}\text{acc}(o \rightarrow n) &= min\left[1, \left\{\frac{(d\,q^{3N} d\,V)_n}{\Omega_{config}(N,\ V,\ U_n)} \frac{1}{N} \frac{(d\,q^3)_o}{(2\delta_{max})^3}\right\} \left\{\frac{(d\,q^{3N} d\,V)_o}{\Omega_{config}(N,\ V,\ U_o)} \frac{1}{N} \frac{(d\,q^3)_n}{(2\delta_{max})^3}\right\}^{-1}\right] \\ &= \frac{\Omega_{config}(N,\ V,\ U_o)}{\Omega_{config}(N,\ V,\ U_n)}\end{aligned} \tag{2.36}$$

The simplification in the expression comes from the fact that the differential elements $d\,V$, $d\,q^{3N}$ and $d\,q^N$ are equal in the old and new state. For this type of move the transition probabilities ($\alpha(o \rightarrow n)$ and $\alpha(o \rightarrow n)$) are equal to each other and cancel out.

The volume change is accompanied with a rescaling of the entire simulation box and the particle positions. The maximum amount of volume change is $\pm\Delta_{max}$ and therefore the new volume can be found within a range of $2\Delta_{max}$ of the old one. While the differential of the particle coordinates ($d\,q^{3N}$) is not the same in the new and old state, that of the reduced coordinates is ($d\,s^{3N} = d\,q^{3N}/V^N$). The acceptance criterion for a volume change can be written as

$$\begin{aligned}\text{acc}(o \rightarrow n) &= min\left[1, \left\{\frac{(d\,s^{3N} d\,V)_n V_n^N}{\Omega_{config}(N,\ V_n,\ U_n)} \frac{(d\,V)_o}{2\Delta_{max}}\right\} \left\{\frac{(d\,s^{3N} d\,V)_o V_o^N}{\Omega_{config}(N,\ V_o,\ U_o)} \frac{(d\,V)_n}{2\Delta_{max}}\right\}^{-1}\right] \\ &= \frac{\Omega_{config}(N,\ V_o,\ U_o)}{\Omega_{config}(N,\ V_n,\ U_n)} \frac{V_n^N}{V_o^N}\end{aligned} \tag{2.37}$$

In this case the simplifications arise as the differential elements $d\,V$ and the reduced coordinates elements $d\,s^{3N}$ are equal in both states. The vapor-liquid phases at coexistence are often separated by several orders of magnitude in density. Thus it is more convenient

to work in the logarithmic scale of volume rather than volume itself. When changes in $\ln V$ are performed, the acceptance criterion is

$$\begin{aligned}
&\mathrm{acc}(o \to n) \\
&= min\left[1, \left\{\frac{(d\, s^{3N} d\ln V)_n V_n^{N+1}}{\Omega_{config}(N,\ \ln V_n,\ U_n)} \frac{(d\ln V)_o}{2\ln\Delta_{max}}\right\}\left\{\frac{(d\, s^{3N} d\ln V)_o V_o^{N+1}}{\Omega_{config}(N,\ \ln V_o,\ U_o)} \frac{(d\ln V)_n}{2\ln\Delta_{max}}\right\}^{-1}\right] \\
&= \frac{\Omega_{config}(N,\ \ln V_o,\ U_o)}{\Omega_{config}(N,\ \ln V_n,\ U_n)} \frac{V_n^{N+1}}{V_o^{N+1}}
\end{aligned} \tag{2.38}$$

In the particle insertion or deletion moves, a particle is added at a random position in the box or one is chosen at random and deleted. In the first case of insertion the acceptance criterion yields

$$\begin{aligned}
\mathrm{acc}(o \to n) &= min\left[1, \left\{\frac{(d\, q^{3N+3} d\, V)_n}{\Omega_{config}(N+1,\ V,\ U_n)} \frac{1}{N+1}\right\}\left\{\frac{(d\, q^{3N} d\, V)_o}{\Omega_{config}(N,\ V,\ U_o)} \frac{(d\, q^3)_n}{V}\right\}^{-1}\right] \\
&= \frac{\Omega_{config}(N,\ V,\ U_o)}{\Omega_{config}(N+1,\ V,\ U_n)} \frac{V}{N+1}
\end{aligned} \tag{2.39}$$

where the new state has one more particle than the old state (from where the term $N+1$ appears in the denominator). For a particle deletion the criterion becomes

$$\begin{aligned}
\mathrm{acc}(o \to n) &= min\left[1, \left\{\frac{(d\, q^{3N-3} d\, V)_n}{\Omega_{config}(N-1,\ V,\ U_n)} \frac{(d\, q^3)_o}{V}\right\}\left\{\frac{(d\, q^{3N} d\, V)_o}{\Omega_{config}(N,\ V,\ U_o)} \frac{1}{N}\right\}^{-1}\right] \\
&= \frac{\Omega_{config}(N,\ V,\ U_o)}{\Omega_{config}(N-1,\ V,\ U_n)} \frac{N}{V}
\end{aligned} \tag{2.40}$$

where the new state has one particle less than the old one.

The simulation of two phases at coexistence for continuum systems should include changes in energy and particle density. In the second case this can be achieved by either performing volume changes or particle insertions and deletions. Therefore first order transitions with the above acceptance criteria for particle displacements and volume changes can be simulated in the isothermal-isobaric ensemble.[12] Those with particle

displacements and insertion/deletion are simulated in the grand canonical ensemble.[13] Since the two changes are independent from each other, the goal is to obtain a two dimensional flat histogram ($H(U,N)$ or $H(U,V)$) between each reduction of the modification factor. At the end, the final density of states dependence on the two other variables can be presented as the excess entropy calculated either as a function of U and N as well as U and $\ln V$ in respectively Figures 2.18 and 2.19. The excess entropy is related to the configurational density of states by the following equation

$$S_{ex} = k_B \ln\left(\frac{N!}{V^N}\Omega_{config}(N,\ V,\ U)\right) \tag{2.41}$$

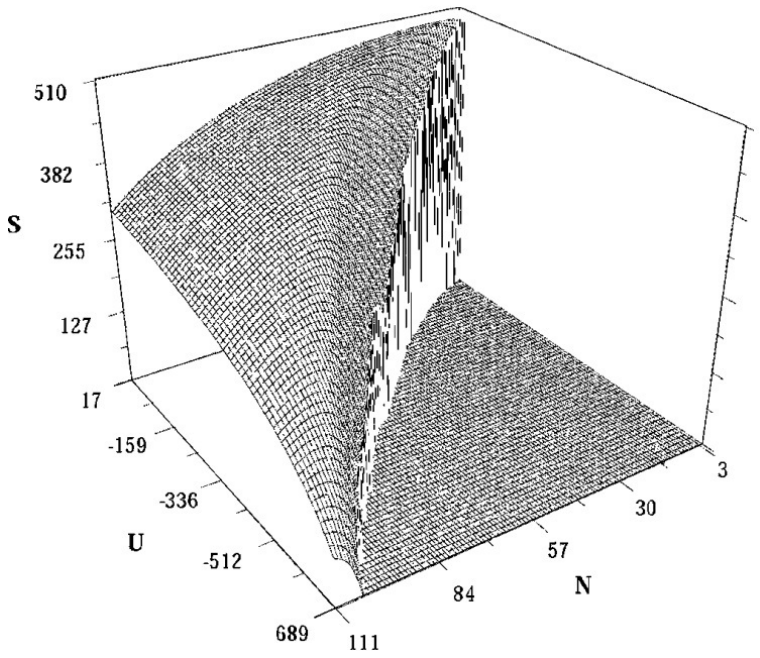


Figure 2.18. Excess free energy as a function of potential energy and number of particles for a Lennard-Jones system.[11]

In Figure 2.18 it can be noticed that the energy range is strongly dependent on the number of particles. At low N, the accessible energy range is rather narrow and a sharp peak is obtained upon decreasing the energy at constant N. Thus a re-establishment of the

energy range is required upon changing the particle number during the simulation. Such a drastic change characteristic for the grand canonical simulation is not present for the isothermal-isobaric ensemble (Figure 2.19). Therefore the volume scaling could be more advantageous in the calculation of the density of states.

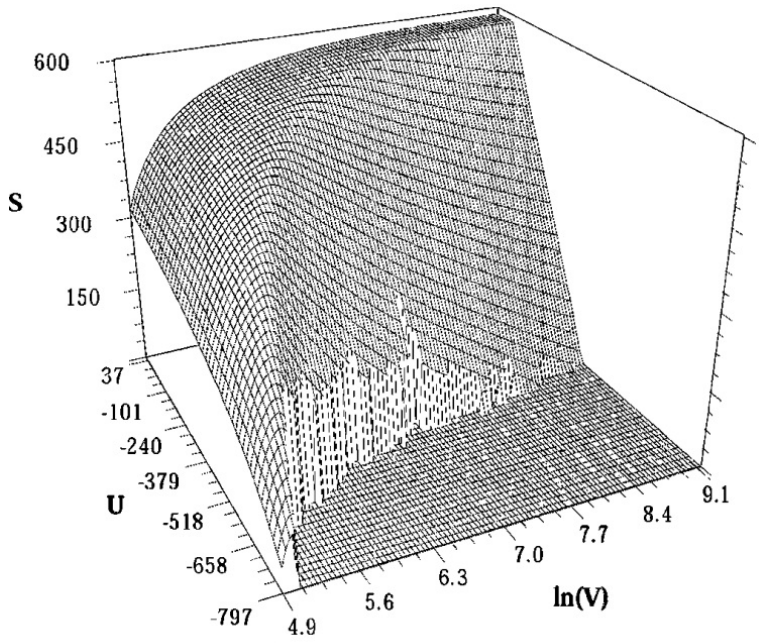


Figure 2.19. Excess free energy as a function of potential energy and volume for a Lennard-Jones system.[11]

The plot of the probability distribution also becomes three dimensional as a function of the energy and number of particles or volume (Figure 2.20).

With the density of states generated after the last flat histogram is obtained, the appropriate partition function (either the grand canonical or isothermal-isobaric one) can be calculated.

$$Q(\mu, V, T) = \sum_{N} \sum_{E} \Omega(N, U) \exp[-(E - \mu N)/k_B T] \tag{2.42}$$

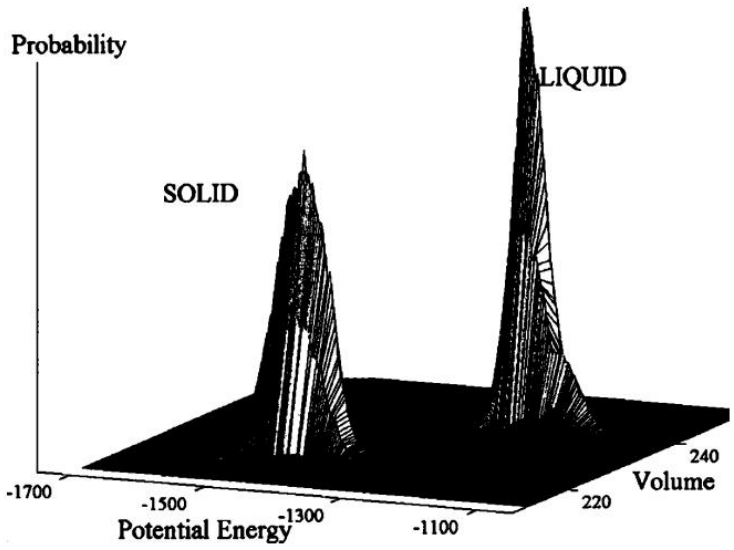


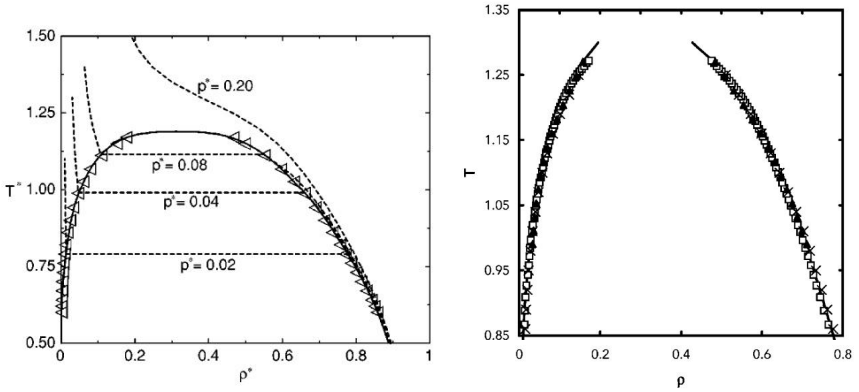
Figure 2.20. Probability distribution for a Lennard-Jones system at a melting point[12]

$$Q(N,P,T) = \sum_{N}\sum_{E} \Omega(V,U)\exp[-(E+PV)/k_BT] \tag{2.43}$$

In the above expressions however the absolute density of states has to be used as opposed to the one coming from the last sweep with the smallest modification factor, known within a multiplicative constant. For these simulations, when $N = 0$ or $V = 0$ the density of states is unity, thus providing a means to determine the absolute value of the partition function.

When the results of the simulation are presented in a two dimensional phase diagrams of the Lennard-Jones system show excellent consistency with experimental data (Figure 2.21(a) and 2.21(b)).

As noted by Yan and coworkers, the fact that the density of states is function of two variables leads to the relevant range of energy being dependent on the other variable (number of particles in that study).[53] In addition, sampling over the energy and number of particles requires a vast amount of CPU time. In order to overcome these issues, the



(a) Phase diagram of the truncated Lennard-Jones fluid. Solid line shows the results of WL simulations in the grand canonical ensemble, while solid line depicts literature data.[53]

(b) Phase diagram of the long-range corrected Lennard-Jones fluid. Squares and triangles correspond to the WL results in respectively the grand canonical and isothermal isobaric ensemble, while crosses are results from histogram reweighting study[11]

Figure 2.21. Vapor-liquid equilibria curves for a Lennard-Jones fluid

authors suggested to perform the simulation at a fixed temperature, so that a flat distribution of solely the number of particles ($H(N)$) is obtained as opposed to $H(U, N)$. The energy distribution of the particle displacement moves is then dictated by the conventional Boltzmann weight

$$\text{acc}(o \rightarrow n) = min\left[1, \exp(-\Delta U/k_B T)\right] \tag{2.44}$$

where $-\Delta U/k_B T = U_n - U_o$ is the energy change associated with the insertion/deletion move.

The probability of any state along the order parameter becomes inversely proportional to the canonical partition function ($Q(N, V, T)$). The insertion and deletion moves are accepted with probabilities, respectively:

$$\mathrm{acc}(o \rightarrow n) = min\left[1, \frac{V}{(N+1)\Lambda^3}\frac{Q(N,V,T)}{Q(N+1,V,T)}\exp(-\Delta U/k_B T)\right] \tag{2.45}$$

and

$$\mathrm{acc}(o \rightarrow n) = min\left[1, \frac{N\Lambda^3}{V}\frac{Q(N,V,T)}{Q(N-1,V,T)}\exp(-\Delta U/k_B T)\right] \tag{2.46}$$

Sampling a first order transition in the grand canonical ensemble by the WL scheme is used in the study of the behavior of alkanes in supercritical carbon dioxide.[13] The carbon dioxide is modeled as a single interaction site with truncated Lennard-Jones potential, while that of hexadecane as five interaction sites connected together by nonlinear elastic springs with the same type of potential between the non-bonded interactions (Figure 2.22).

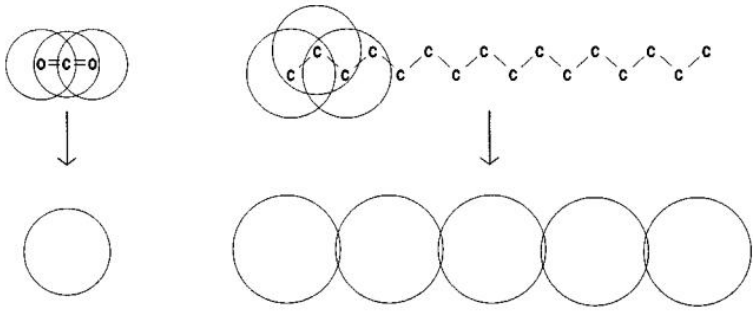


Figure 2.22. Carbon dioxide modeled as a single Lennard-Jones particle and hexadecane modeled as a chain of 5 hexadecane particles bonded together.[13]

Since the simulation is performed in the grand canonical ensemble with long molecules, the insertion of particles in the dense liquid region leads to particle overlaps. In order to overcome this problem, a configurational bias Monte Carlo scheme is implemented for the insertion and deletion moves. The multicanonical method developed by Berg and Neuhaus[47] is used in order to sample the density of the liquid and vapor phase uniformly at a given chemical potential and temperature, while the histogram reweighting

is used to find the probability distribution at another chemical potential and temperature.

The range of densities comprising the two phases at coexistence is sampled by varying the number of particles (N) and the uniformity in states along the order parameter is achieved through the introduction of the weight function $w(n)$. A biased probability $p_{bias}(N) = p(N)\exp(-w(N))$ is introduced in such a way that it becomes constant for the choice of $w(N) = \ln p(N)$. The WL scheme is implemented in the estimation of $w(N)$ in a similar fashion the density of states is found in the original WL method. An initial value of $w(N) = 0$ is given for the entire density range. Each sampling of a new state is accompanied with a modification of the weight function with an increment δw. A histogram of visited states ($H(N)$) is also updated at each attempted move. When this histogram is sufficiently flat ($|H(N) - \langle H(N)\rangle| < 0.5\,\langle H(N)\rangle$ for every N), the increment is decreased by a factor of 10 and the random walk in N is repeated after the histogram is reset to zero. The initial value of δw is set to 10^{-3} and after three iterations a working guess for $w(N)$ is obtained (Figure 2.23). The initial increment value δw and the flatness criterion depends on the system size and in order to avoid errors the multicanonical sampling (δw) is encouraged to be performed for the production run.

The difficulty of obtaining a two dimensional flat histogram of visited states is to some extent attenuated by combining the WL scheme with a parallel tempering technique in the simulation of folding processes in proteins by Rathore *et al.*[14,54] Unlike previous simulations where long chain molecules are simulated using a lattice model, the authors manage to implement the WL scheme to study the folding of chain proteins in continuum. In such systems, the energy walk happens in a much larger energy space and there is higher degeneracy of the energy levels than the lattice systems. Therefore the random

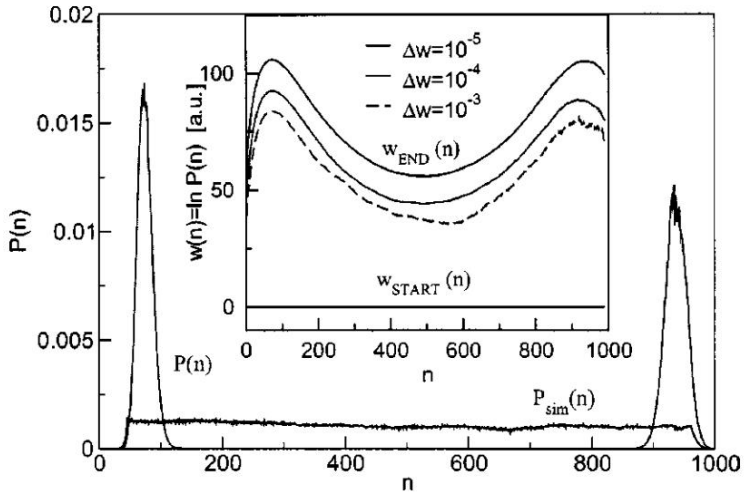


Figure 2.23. Probability distribution as a function of the number of particle and convergence of the weight function (inset).[13]

walk is coupled with a parallel tempering methodology. Two different protein molecules are simulated *in vacuo* and in an implicit solvent. The protein molecule is modeled with the CHARMM19 force field with bonding interactions taking into account bond stretching, angle banding and proper and improper dihedral changes, while the non-bonding interactions are modeled with the Lennard-Jones and Coulomb potentials. The latter type of interactions occurs between the sites on the same molecule which are separated by at least three successive bonds.

In the proposed algorithm, only conformational changes of the protein molecule are investigated. Instead of using the conventional MC displacement moves, the authors suggested to use the hybrid Monte Carlo scheme. The MC move consists of performing several molecular dynamics steps *m* in the NVE ensemble. The velocities are assigned according to a Gaussian distribution corresponding to the temperature of the simulation box. Nevertheless a random walk is performed in the conformational phase space and no real dynamic behavior is traced. The acceptance criteria given by Equation 2.7 is modified

to include the total kinetic energy K

$$\mathrm{acc}(o \rightarrow n) = \min\left[1, \frac{\Omega(U_o)}{\Omega(U_n)}\exp(-\Delta K/k_B T)\right] \tag{2.47}$$

where ΔK is the change in total kinetic energy upon changing the system from an old to a new state. In order to enhance the configurational rearrangement, another type of move is performed - the pivot move. A rotation angle is selected in a certain interval $0 < \theta < \theta_{max}$ and the rotation is effectuated around a randomly oriented axis passing through either carboxyl or amino segments of the molecule. The move is accepted with the probability given in Equation 2.7.

The energy range comprising the folded and unfolded protein molecule conformations and the number of possible states for a continuum system are much larger than for a lattice one. Therefore the convergence of the simulation slows down as the sampled energy range is increased. If however the energy range is subdivided into smaller intervals, one risks to get trapped in a particular conformational well of the protein. This is not desirable since some of the moves can lead to a drastic conformational change and thus the protein molecule cannot be restructured. The solution proposed by the authors is to create several replicas of the simulation box, each at a different temperature. The replicas do not interact with each other and for each of them an energy interval is assigned. When a replica with higher energy is chosen, the maximal rotation angle (θ_{max}) related to the pivot move is increased.

The density of states estimation is performed independently in these boxes. The conformations in different replicas are swapped on a regular basis. The swapping move is accepted if the potential energies of both replicas lie within the overlapping region of the

energy intervals for the two simulation boxes. The idea is to keep narrow energy intervals and small θ_{max} in the folded regime in order to keep a copy of the near-folded conformations. Narrow energy windows are necessary in the transition region as well in order to achieve faster convergence. On the other side of the overall energy range, the intervals ranges as well as θ_{max} are larger so that major rearrangements of the proteins are effectuated and attempted replica exchange moves are accepted. The heat capacity is calculated from Equation 2.16 and plotted as a function of temperature in Figure 2.24. In the case of Met-enkephalin similar transition temperatures corresponding to the peak of the curve for the two types of implicit solvents (an implicit solvent model based on the solvent access surface area (SASA) and a solvent effect model based on the distance-dependent dielectric (DDE) function $\varepsilon(r) = 2r$) is obtained. The two transition temperatures are also comparable to the transition temperature region upon tracing the average end to end distance ($\langle d_{e-e} \rangle_T$) between the N- and C-terminus of the protein molecule (Figure 2.25). The protein molecule is extended at large temperatures and folded at lower temperatures which is confirmed experimentally.

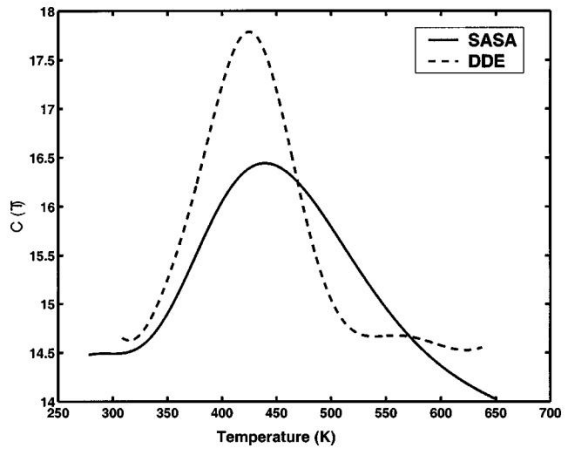


Figure 2.24. Specific heat capacity per unit residue as a function of temperature for Met-enkephalin.[14]

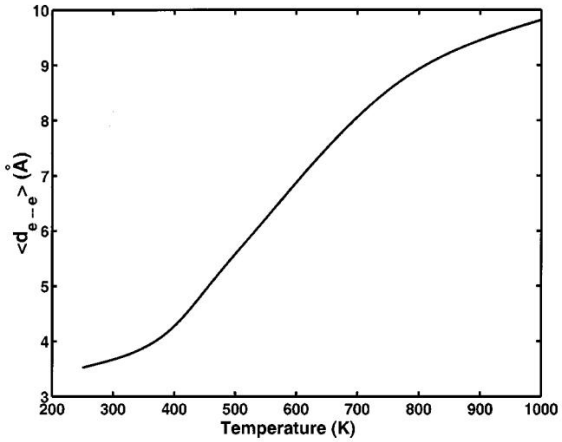


Figure 2.25. Average end to end distance $\langle d_{e-e} \rangle_T$ as a function of temperature for Met-enkephalin.[14]

2.3 Vapor-liquid phase equilibria

The idea by Yan *et al.* of performing the simulation at constant temperature was later developed by by Ganzenmüller and Camp,[45] who developed the methodology in the canonical, isothermal-isobaric and grand canonical ensembles. Sampling along the order parameter at a constant temperature allows one to alleviate the issues related to the estimation of density of states in previous simulations of continuum systems.

In any given ensemble, the choice of the biasing function that is dynamically updated is dictated by the need to achieve a uniform sampling of an extensive variable which allows the sampling in both vapor and liquid phases. Let this variable be called X. The probability of finding X is equal to:

$$p(X) = \int p(\Gamma, X) d\Gamma \tag{2.48}$$

where Γ is a specific configuration and $\Pi(\Gamma, X)$ is the joint probability of being in this

specific configuration with a specific value of X. The biased distribution, $\Pi_{bias}(X)$ should be uniform - the probability of finding any of the states along the region comprising the vapor and the liquid state is equal. If we choose the joint biased probability in the following manner:

$$p_{bias}(\Gamma, X) = \frac{p(\Gamma, X)}{p(X)} \tag{2.49}$$

then $p_{bias}(X)$ becomes one for any value of X:

$$p_{bias}(X) = \int p_{bias}(\Gamma, X) d\Gamma = \int \frac{p(\Gamma, X)}{p(X)} d\Gamma = 1 \tag{2.50}$$

The detailed balance equation is then:

$$p_{bias}(\Gamma_o, X_o) acc_{(o \to n)} = p_{bias}(\Gamma_n, X_n) acc_{(n \to o)} \tag{2.51}$$

where $acc(o \to n)$ is the acceptance probability of a trial move from an old (o) to a new (n) configuration. The Metropolis solution becomes:

$$acc_{(o \to n)} = min \left[1, \frac{p_{bias}(\Gamma_n, X_n)}{p_{bias}(\Gamma_o, X_o)} \right] \tag{2.52}$$

2.3.1 Sampling in the canonical ensemble

The WL method was developed as an improvement of the conventional MC algorithm for sampling only the energy space in a lattice system. The analogue sampling in continuum systems would be performed in the canonical ensemble. The biased distribution is derived from the Boltzmann probability distribution in Γ:

$$p(\Gamma) = \frac{V^N \exp[-\beta U(\Gamma)]}{N!\,\Lambda^{3N} Q(N,V,T)} \tag{2.53}$$

where $\Lambda = \sqrt{h^2/2\pi m k_B T}$ is the de Broglie thermal wavelength and $U(\Gamma)$ is the potential energy. Since in the canonical ensemble the only variable is $U(\Gamma)$ (function of the particle positions), $p(\Gamma)$ can be considered as the joint distribution. The energy distribution should include those configurations with energy E:

$$p(E) = \int p(\Gamma)\delta(U(\Gamma) - E)\mathrm{d}\Gamma \tag{2.54}$$

where δ is the Dirac delta function. The density of states, with energy E is given by:

$$\Omega(N,V,E) = \int \delta(U(\Gamma) - E)\mathrm{d}\Gamma \tag{2.55}$$

and hence the energy distribution can be replaced with the following expression:

$$p(E) = \Omega(N,V,E)\frac{V^N \exp[-\beta E(\Gamma)]}{N!\,\Lambda^{3N} Q(N,V,T)} \tag{2.56}$$

The biased probability is obtained as shown in Equation 2.49:

$$p_{bias}(\Gamma, E) = \frac{1}{\Omega(N,V,E)} \tag{2.57}$$

With this biased probability the final expression of the acceptance criterion can be derived, according to the detailed balance condition 2.51:

$$acc_{on} = \min\left[1, \frac{\Omega(N,V,E_o)}{\Omega(N,V,E_n)}\right] \tag{2.58}$$

The biasing function in the canonical ensemble thus becomes the density of potential energy states with N and V fixed. The Boltzmann probability distribution $p(E)$ is found by performing a random walk which would result in a flat histogram since the acceptance probability is biased according to Equation 2.51. $\Omega(N, V, E)$ is modified in a systematic way so that the random walk over the allowed energy range would converge to the true value.

2.3.2 Sampling in the isothermal-isobaric ensemble

In the isothermal isobaric ensemble number of particles (N), pressure (P) and temperature (T) are fixed. The extensive variable that allows the sampling of the vapor-liquid region is the volume (V). In this ensemble the joint distribution is given by:

$$p(V, \Gamma) = \frac{V^N exp[-\beta U(\Gamma) - \beta PV]}{N!\,\Lambda^{3N} Q(N, P, T)} \tag{2.59}$$

where $Q(N, P, T) = \int_0^\infty Q(N, V, T) exp(-\beta PV) dV$ is the isothermal isobaric partition function and $Q(N, V, T)$ is the canonical partition function. The integration of the joint probability over the possible configurations then yields:

$$p(V) = \frac{Q(N, V, T) exp(-\beta PV)}{Q(N, P, T)} \tag{2.60}$$

Following the expression for the joint biased probability in equation 2.49, in the isothermal isobaric ensemble becomes:

$$p_{bias}(V, \Gamma) = \frac{V^N exp[-\beta U(\Gamma)]}{N!\,\Lambda^{3N} Q(N, V, T)} \tag{2.61}$$

which leads to an acceptance probability:

$$acc_{(o\rightarrow n)} = min\left[1, \frac{Q(NV_oT)V_n^N exp[-\beta U(\Gamma_n)]}{Q(NV_nT)V_o^N exp[-\beta U(\Gamma_o)]}\right] \tag{2.62}$$

In this expression the variable that is unknown *a priori* and plays the role of the biasing function is the canonical partition function $Q(N, V, T)$ and obtaining a good estimation of it is at the core of the simulation. Another important characteristic of the WL method in the NPT ensemble is that the pressure does not figure in the acceptance rule of equation 2.62. Thus the coexistence pressure is found after the simulation converges and does not have to be predetermined in advance.

In the classical Wang-Landau method of spherical particles in the N, P, T ensemble, two types of moves are performed, random particle translation moves and volume changes. For this purpose, the volume domain comprising the volume and liquid range is divided in equal intervals in lnV. The simulation starts by attributing an arbitrary value to $Q(N, V, T) = 1$. Each time a new volume interval is sampled, it is updated by a convergence factor $f(Q(NV_oT)) \rightarrow fQ(NV_oT)$, starting with $lnf = 1$. The updated partition function is higher in value and, thus, as it can be seen from equation 2.62, the acceptance probability of visiting a new volume increases. This constant updating is the driving force behind the uniform sampling. Not only it allows to visit low probability volumes, but also the partition function converges towards its true value. Once all states in each volume interval are sampled at least a certain number of times, which ensures a reasonably flat sampling, another sweep is run with a lower convergence factor ($f_n = \sqrt{f_{n-1}}$). It has to be stressed that the convergence of the simulation is monitored by tracing the histogram of visited states and the running estimate of the partition function.

The refinement of the convergence factor is necessary, as a better estimate of $Q(N, V, T)$ is obtained after each sweep. Even though after it the random sampling of the volume interval starts all over, the estimate of the partition function is kept in the computer memory. The refinement of the convergence factor f is accompanied by a refinement in $Q(N, V, T)$ and when it has an allure of smooth and continuous function, the simulation is over.

With the formalism developed by Ganzenmüller and Camp the WL method is applied to study spherical particles such as Lennard-Jpnes fluid and charged soft spheres.[45] The canonical partition function is used to calculate the densities at equilibrium at the temperatures at which the runs are performed. First the pressure at coexistence can be estimated by calculating the probabilities associated with the liquid (p_{liq}) and vapor (p_{vap}) phases.

$$p_{liq} = \int_0^{V_b} \Pi(V)dV \quad \text{and} \quad p_{vap} = \int_{V_b}^{\infty} \Pi(V)dV \tag{2.63}$$

where V_b correspond to the volume with minimum probability, situated between the vapor and liquid phase, and calculated with Equation 2.60. The pressure in this expression (2.60) is changed until the two probabilities p_{liq} and p_{vap} are equal, at which point the pressure at coexistence is obtained. With it, the densities at equilibrium can be calculated. First, the vapor density is given by

$$\rho_{vap} = \frac{\int_{V_b}^{\infty} (N/V)V^N Q(N, V, T) exp\left(-\beta P_{eq} V\right) dV}{\int_{V_b}^{\infty} Q(N, V, T) exp\left(-\beta P_{eq} V\right) dV} \tag{2.64}$$

and then the liquid density is estimated

$$\rho_{liq} = \frac{\int_0^{V_b} (N/V) V^N Q(N,V,T) exp\left(-\beta P_{coex} V\right) dV}{\int_0^{V_b} Q(N,V,T) exp\left(-\beta P_{coex} V\right) dV} \tag{2.65}$$

With these data the phase diagram of the Lennard-Jones fluid is established and excellent consistencies between results from the WL scheme and the transition matrix MC method are obtained with maximum deviation between the two sets of data of 1.0 % and avergae deviation of 0.4 %.

While largely studied on model fluids consisting of structureless spherical particles, the implementation of the WL scheme on molecular fluids was first described by Desgranges and Delhommelle.[16] The method is applied to rigid molecules such as benzene and flexible chains such as n-alkanes. The authors successfully implemented the hybrid Monte Carlo scheme in the WL method (HMC-WL) in order to sample efficiently the liquid states of the system. Thus the acceptance criterion for the updated particle positions after each HMC move becomes:

$$acc(o \rightarrow n) = min\left[1, exp\left[-\beta(E_n - E_o)\right]\right] \tag{2.66}$$

where E is the total energy of the system.

The HMC-WL algorithm is applied to a Lennard-Jones fluid for comparison reasons. The volume range is divided into equal portions (bins) of $\ln V$ and the accuracy of the results is tested by comparing the simulation data for different number of bins to the TM data obtained by Errington[15] (Figure 2.26). With the 100 bins division of the volume range the HMC-WL simulation data tends to overestimate the TM results for liquid

densities at coexistence by 1 to 2 % at low temperatures. The overestimation is similar to that observed by Ganzenmüller and Camp.[45] The accuracy can be improved by increasing the number of bins as seen in Figure 2.26. It decreases to less than 0.5 % when the number of bins reaches 400. The slight improvement in accuracy is achieved at the expense of computational time, which is more than double for passing from a system of 150 to 400 bins.

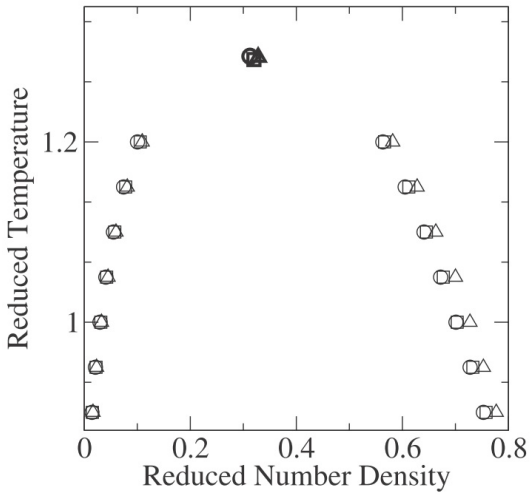


Figure 2.26. Vapor-liquid phase diagram of Lennard-Jones fluid. Data from the TM method by Errington[15] are denoted as open circles, while results from the HMC-WL simulation by open triangles for the volume range divided into 100 bins and with open squares for the volume range divided into 400 bins.[16]

The results obtained for benzene and n-alkanes show a very good consistency with experimental data in terms of the saturation vapor pressure (Figure 2.27) and vapor-liquid equilibria curves (Figure 2.28). The accuracy of the simulation results proves the reliability of the method upon simulating molecular fluids with different architectural structures.

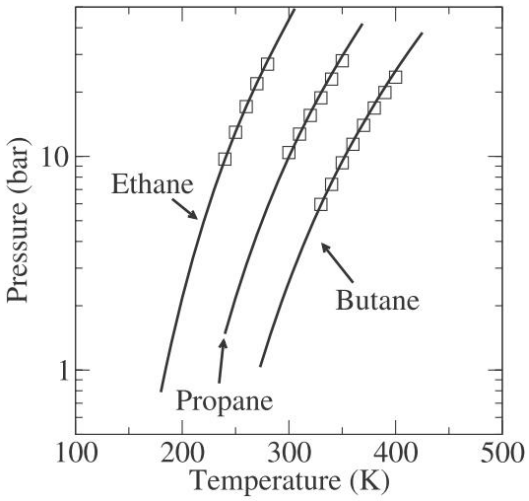


Figure 2.27. Saturation pressure obtained with the HMC-WL method: simulation results (open squares) and experimental data (solid line)[16]

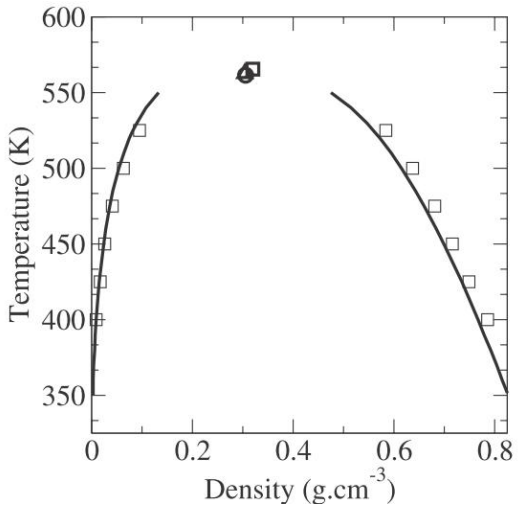


Figure 2.28. Vapor-liquid equilibria curves obtained with the HMC-WL method (open squares) and experimental data (solid line).[16]

CHAPTER III

VAPOR-LIQUID EQUILIBRIA OF COPPER

Molecular simulations on vapor-liquid equilibria of metals are very limited. Data is available only for a few elements; among them, aluminum, gold, potassium, sodium, copper, nickel and mercury.[55–58] Conventional Molecular Dynamics (MD) simulations have been used in the case of nickel and mercury, which however cannot directly estimate the properties at equilibrium. MD simulations permit the estimation of thermodynamic properties in either vapor or liquid phase, but not both phases at coexistence. This can be explained by the high energy barrier between the two phases, which acts as a bottleneck that prevents the sampling of the density range of interest. In order to find the densities at coexistence of the vapor and liquid phases, the authors have applied an empirical equation of state, that completes the MD data.[57,58] The accuracy of the results in these two cases nonetheless hinges on not only the choice of the force field, but also the equation of state.

We propose a combination of the Wang Landau method in combination with the hybrid Monte Carlo technique in order to study the phase behavior of copper in the isothermal isobaric ensemble. Thus the difficulty related to the insertion of particles in the dense liquid phase of the GEMC method is solved. Our method does not rely on a configurational bias related to the insertion move, since sampling is performed in the NPT ensemble, where changes in the volume allow to reach states related to the vapor and liquid phase as well as the interfacial region.

3.1 Potential model

The choice of the force field is crucial for the accuracy of the simulated properties. Vapor-liquid equilibria curves of metals have been simulated by using two types of force fields - an embedded atom potential (EAM) and a Morse pair potential.

The Morse pair potential contains only two body interaction terms and has the following form:

$$U = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} D\left(exp[-2\alpha(r_{ij}-r_0)] - 2exp[-\alpha(r_{ij}-r_0)]\right) \tag{3.1}$$

where D is the dissociation energy, α is a constant with dimensions of reciprocal length and r_0 is the equilibrium distance of the two particles. This form has been used by Singh and coworkers for the simulation of gold, potassium, sodium and copper[56] as well as by Cheng and Xu in the MD computation of nickel.[57] The potential allows faster computation of the potential energy of the system, as there are no many body terms. The accuracy of the potential however is questioned, as the models used are parameterized for metals in the solid state and the local density is not taken into account.

An embedded atom potential can be presented in the following form:

$$U = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\phi(r_{ij}) + \sum_{i}^{N} F(\rho_i) \tag{3.2}$$

where $\phi(r_{ij})$ is a two body interaction term and $F(\rho_i)$ is a many body term, with ρ_i being the local density associated with atom i. The latter term is an important part of simulation of metals as it accounts for the strong cohesive forces. It is therefore the potential of choice for the simulation of aluminum by Siepmann and cowerkers.[55] Another form of the

embedded atom potential, the Sutton-Chen EAM, has also proven its ability to accurately predict rheology and crystallization processes in liquid metals[59,60] and therefore is the model of choice in our simulations. The two body term in Equation 3.2 is given by

$$\phi(r_{ij}) = \varepsilon\left(\frac{a}{r_{ij}}\right)^{n} \tag{3.3}$$

and the many-body term by

$$F(\rho_i) = -\varepsilon c\left[\sum_{j\neq i}\left(\frac{a}{r_{ij}}\right)^{m}\right]^{0.5} \tag{3.4}$$

The cutoff distance is set to be twice the lattice parameter a. All the parameters are taken from the work of Luo *et al.*[25] and presented in Table 3.1.

Table 3.1. qSC-EAM potential parameters for copper.[25]

a (Å)	ε (10^{-2} eV)	c	m	n
3.6030	0.57921	84.843	5	10

3.2 Hybrid Monte Carlo Wang-Landau simulations

In the formalism of the Wang-Landau scheme at a constant temperature, the conventional MC displacement moves are used to explore the phase space.[45] In our simulation of copper the hybrid Monte Carlo scheme is used for the random displacement moves. The acceptance criterion for a particle displacement move is given in Equation 2.66. The velocity-Verlet algorithm is used to integrate the equations of motion, due to its area preserving properties and time reversibility. The advantage of using HMC over the MC scheme is twofold. First, the MD trajectories account for the global and concerted

moves of the particles. The moves are not completely randomized, but deterministic in nature, which in the case of dense liquids such as metals, provides higher acceptance rates than when using an MC scheme. Second, the time step of the MD move can also be set higher than that of a conventional MD simulation, as some of the trajectories are allowed to be rejected. Long MD trajectories result in an efficient sampling of the system.[16]

The simulation of copper is divided into two parts. In the first part, the goal is to find the vapor-liquid equilibrium curve and critical properties. The temperature range is therefore chosen to be close, yet below the critical temperature. In the second part, the simulation aims to estimate the boiling point of the fluid and therefore the runs are performed at temperatures close to the experimental boiling point.

For the first part, random volume changes and HMC moves are performed at constant temperature, starting with 4800 K until the simulation converges. The other runs are performed with temperature increments of 100 K, ending the simulation with a run at 4100 K. For this lowest temperature, the upper and lower bounds of the sampled volume are chosen so that the maximum distribution probabilities of the densities of two phases are within these bounds. The upper and lower bounds for the volume are chosen to correspond to densities of 0.01 and 6.5 g cm^{-3}. The simulated systems consist of 200 atoms. In the case of a metal for which these bounds are unknown, appropriate values can be attributed by running a short HMC WL simulation with just a few sweeps (few refinements of the convergence factor). The density range is translated in a volume range divided into 150 intervals of equal lengths in $\ln V$.

For the second part, the temperature range is between 2900 to 3000 K. The upper and lower bounds for the volume are chosen to correspond to densities of 0.0001 and 7

g cm^{-3}. The increase in the density range is due to the pressure drop. The simulated systems consist of 200 atoms. The density range is translated in a volume range divided into 400 intervals of equal lengths in $\ln V$.

Once the canonical partition function at the chosen temperatures is obtained, the isothermal isobaric partition function can be calculated, which in turn can give rise to densities in the liquid and vapor phases.

$$Q(NPT) = \int_0^\infty Q(NVT) exp\left(-\beta PV\right) dV \tag{3.5}$$

The pressure at coexistence can be estimated by calculating the probabilities associated with the liquid (p_{liq}) and vapor (p_{vap}) phases.

$$p_{liq} = \int_0^{V_b} \Pi(V) dV \quad \text{and} \quad p_{vap} = \int_{V_b}^\infty \Pi(V) dV \tag{3.6}$$

where V_b corresponds to the volume with minimum probability, situated between the vapor and liquid phase, and calculated with Equation 2.60. The pressure in this expression (2.60) is changed until the two probabilities p_{liq} and p_{vap} are equal, at which point the pressure at coexistence is obtained. Thus densities at equilibrium can be calculated:

$$\rho_{vap} = \frac{\int_{V_b}^\infty (N/V) V^N Q(N,V,T) exp\left(-\beta P_{eq} V\right) dV}{\int_{V_b}^\infty Q(N,V,T) exp\left(-\beta P_{eq} V\right) dV} \tag{3.7}$$

$$\rho_{liq} = \frac{\int_0^{V_b} (N/V) V^N Q(N,V,T) exp\left(-\beta P_{coex} V\right) dV}{\int_0^{V_b} Q(N,V,T) exp\left(-\beta P_{coex} V\right) dV} \tag{3.8}$$

A phase diagram of the vapor and liquid regions for a given temperature and

density can be constructed by connecting the points corresponding to the vapor and liquid densities.

3.3 Evaluation of the critical point

Due to the large degree of fluctuations that occur near the critical point, direct estimation is unreliable. The estimation of the critical point in this work is achieved first by evaluating the critical temperature T_c, with a density scaling law:

$$\rho_{liq} - \rho_{vap} = \mathrm{B}\,(T_c - T)^{\beta} \tag{3.9}$$

where B and T_c are obtained after least square fitting of the available data of ρ_{liq} and ρ_{vap} in the above equation and $\beta = 0.3265$ is the 3D Ising order-parameter exponent. The critical density ρ_c is then estimated using the law of rectilinear diameters given by

$$\frac{\rho_{liq} + \rho_{vap}}{2} = \rho_c + \mathrm{A}\,(T - Tc) \tag{3.10}$$

3.4 Results and discussion

3.4.1 Vapor-liquid equilibria and critical point

The convergence of the simulation, following a WL scheme, is assessed by tracing the evolution of the biasing canonical partition function throughout the simulation. The flatness of the histogram of visited states is governed by the biased acceptance criterion, which is a function of $Q(N, V, T)$. The histogram is considered to be reasonably flat if each volume interval is visited at least 500 times. Since $Q(N, V, T)$ is continuously updated, it is important to trace its evolution with volume for each sweep, in order to accurately evaluate its convergence. The plotted running estimate of $Q(N, V, T)$ is shown in Figure

3.1. It becomes more and more accurate as the convergence factor f decreases. The roughness of the curve gradually diminishes as the simulation progresses until f becomes close to 1 ($lnf = 10^{-5}$). At this convergence factor, the smoothness of the curve indicates a good final estimate of the biasing function. Further updates would not change significantly the outcome of the simulation since the running estimate of the partition function would stay close to that in the preceding sweep.

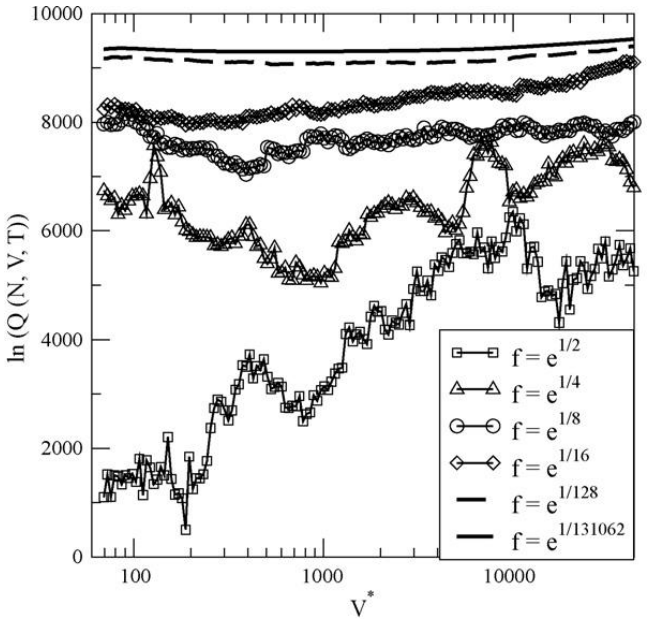


Figure 3.1. Canonical partition function $Q(N, V, T)$ as a function the reduced volume for different convergence factors f.[17]

For each temperature run, $Q(N, V, T)$ is set to 1 and allowed to evolve. Thus the outcome of each run is independent of the preceding one. The same qualitative behavior, however, is obtained at each temperature as shown in Figure 3.2. This proves, once again, that the simulation runs have converged and an accurate estimates of $Q(N, V, T)$ is reached for every run.

With an accurate estimate of the partition functions, the coexistence properties of copper can be represented graphically. In Figure 3.3, the liquid and vapor equilibrium

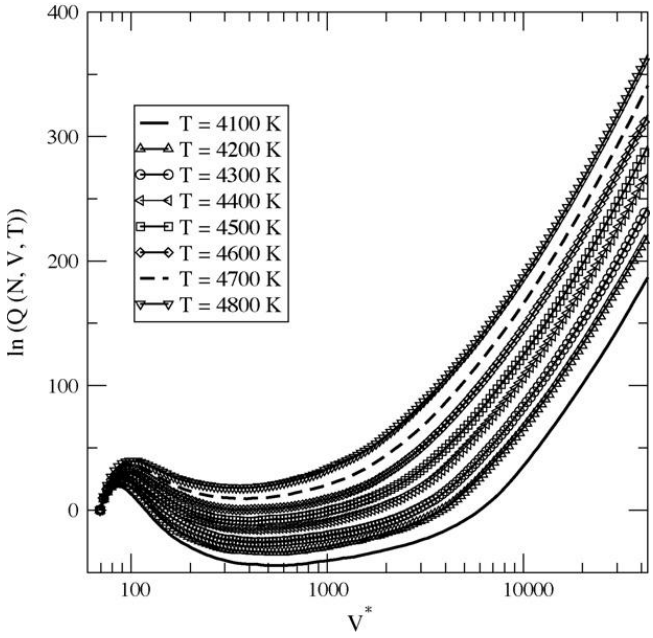


Figure 3.2. Qualitative behavior of lnQ(N, V, T) as a function of the reduced volume for different temperatures used in this work.[17]

densities are plotted at each temperature, using Equations 3.8 and 3.7, respectively. The open square represents the critical point obtained by using the density scaling law (Equation 3.9) and the law of rectilinear diameters (Equation 3.10). The critical point is well above the highest temperature at which the simulation is performed, in order to avoid any finite-size effects. These effects appear when the correlation length of the system exceeds the size of the system. In the vicinity of the critical point, the divergence of the correlation length cannot be captured by the size of the simulation box and therefore it is essential to avoid sampling this region.[61] The critical temperature and density were found to be, respectively, 5696±50 K and 1.80±0.01 g/cm^3.

Critical pressure is obtained by fitting the coexistence pressures, using Antoine's law:

$$\log_{10} P_{coex} = A - \frac{B}{T + C} \tag{3.11}$$

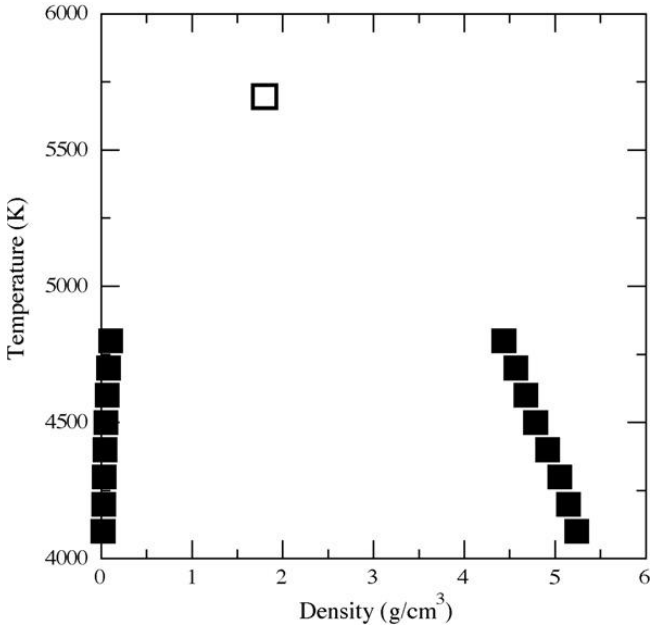


Figure 3.3. Vapor-liquid equilibria for Cu obtained by NPT HMC-WL simulations (the statistical uncertainty is less than the symbol size). The open square represents the critical point extrapolated from the set of results.[17]

where P_{coex} is in Pa, T is in C and A, B and C are fitting parameters. Using the equilibrium pressures from our simulation and plugging the in Equation 3.11, a critical pressure of 1141±100 bar is finally obtained.

Critical and vapor-liquid equilibrium properties of copper available in the literature are very scarce. Saturation pressures at such elevated temperature cannot be found in literature, but at lower temperature data is available.[18] On the plot in Figure 3.4 the experimental data is plotted (circles) and extrapolated (solid line), using Antoine's law (Equation 3.11). The plot shows that extrapolation of the experimental set lies closely to the simulation data at elevated temperatures, which confirms the consistency between the experimental and HMC-WL results.

Finally, the critical properties from the HMC-WL simulations are compared to experimental data and another simulation work (Table 3.2). The simulation results

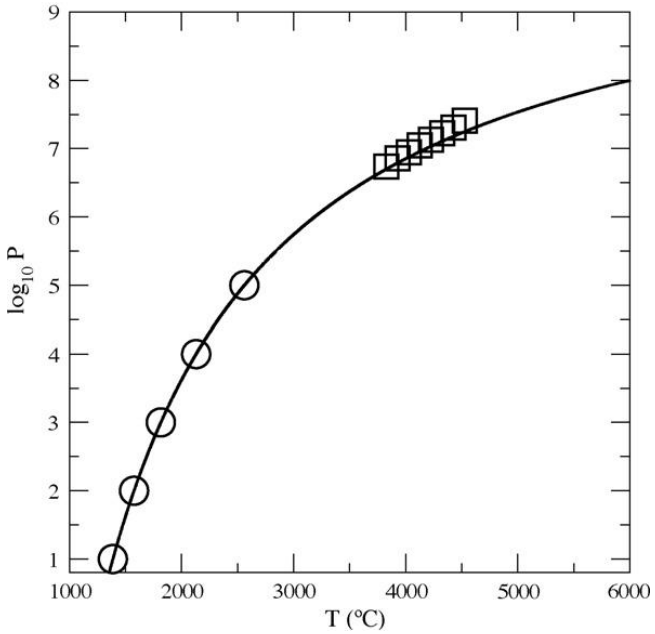


Figure 3.4. Saturation vapor pressure in function of temperature obtained from experiments (open circles),[18] extrapolation with Antoine's law of these experimental results (solid line) and our simulations (open squares).

presented in the second row of the table are given by Singh *et al.*[56] The method adopted by the authors is transition matrix Monte Carlo simulation in the grand canonical ensemble with a two-body Morse potential, used as a force field. Experimental critical properties differ significantly one from another. T_c can be roughly found in the region of 5000 to 8000 K, while the critical pressure varies even more depending on the source with an order of magnitude between the values of approximately 500 and 6000 bar.[24,62] These large discrepancies can be attributed mainly to the difficulty of performing an experiment in such harsh conditions of elevated temperatures and pressures. The results from our simulations lie within the experimental range of critical temperature and pressure. Furthermore, our critical density is very close to the critical density obtained experimentally by Hess.[62] The data obtained from Singh *et al.*,[56] on the other hand, overestimate all of the critical properties presented in Table 3.2. The largest critical

temperature is overestimated by 12 %, while the largest critical pressure and density by 64% and 36% respectively. These results can be attributed most probably to the two-body Morse potential used in their simulations. While the simulation converges faster with this type of force fields and it has shown some success in modeling liquid metals, it does not contain the many body term that accounts for the strong cohesive forces in metals, such as in the qSC-EAM.

Table 3.2. Critical properties of Cu.

	T_c (K)	P_c (bar)	ρ_c (g/cm^3)
This work	5696 ± 50	1141 ± 100	1.80 ± 0.03
Sim.[56]	8650 ± 50	9543 ± 180	2.631 ± 0.003
Exp.[62]	7696	5829	1.930
Exp.[24]	5140-5580	420-600	–

3.4.2 Boiling point

The formalism for the VLE of copper is also used to determine the saturation pressure at lower temperatures. Upon obtaining the canonical partition function, the probability distribution is found using Equation 2.60 and plotted in Figure 3.5. The areas under the two peaks corresponding to the liquid (left) and vapor (right) phases are integrated. When they are equal, the pressure at coexistence P_{coex} is found (Equation 3.6).

The saturation pressures are plotted as a function of temperature in Figure 3.6, combining the results at high (open squares) and low (open circles) temperatures. The results are consistent as shown by the close fit to Antoine's relation (Equation 3.11). It allows us to estimate the standard boiling point of copper at T_b = 2936 ± 6 K. This estimate is in an excellent agreement with simulation results from a constant-NPH adiabatic GEMC simulation, with a standard boiling point of T_b = 2928 ± 4 K.[20] Both

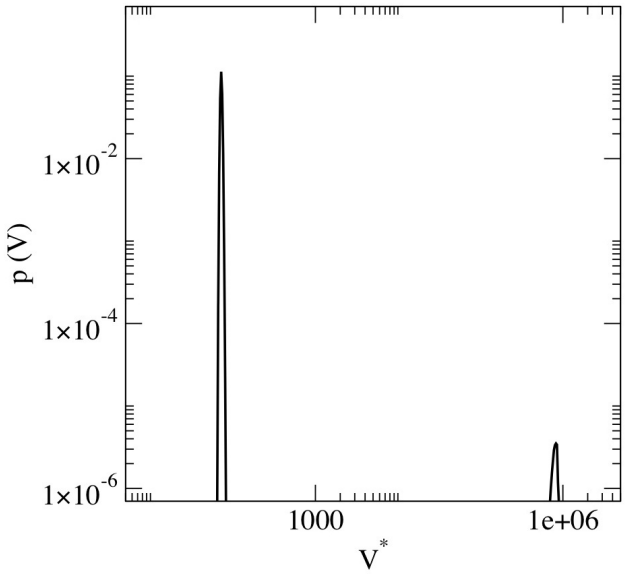


Figure 3.5. Volume probability distribution for copper at 3000 K at coexistence.

results are in a reasonable agreement with the experimental value of T_b^{exp} = 2835 K.[19] The simulation results from our work and those using the GEMC technique overestimate the boiling point by roughly 100 K. As noted by Gelb and Chakraborty, this could be due to the inaccuracy in the force field when applied to the vapor phase.[20]

3.5 Technical details

3.5.1 Molecular dynamics

Sampling of the phase space through a molecular dynamics simulation is governed by the Newton's equations of motion in a classical system of particles. The equations of motions are solved for all the particles to a certain degree of accuracy in a discrete step-by-step manner along time.

In molecular dynamics simulations the average of a macroscopic observable is obtained after letting the system of N particles to evolve in time and therefore it is necessary to trace its evolution. Lagrangian mechanics can be used to study the trajectory

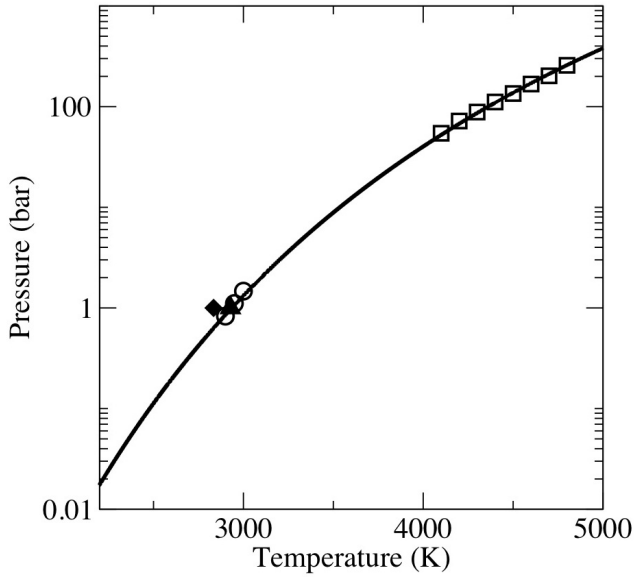


Figure 3.6. Saturation pressure of copper as function of the temperature at coexistence. Low temperature HMC-WL results (open circles), high temperature HMC-WL results (open squares), experimental boiling point (filled diamond),[19] and constant-NPH adiabatic GEMC simulations (filled triangle).[20] The solid line is fit of our simulation data at high and low temperatures using Antoine's law.

of a system of particles. The Lagrangian is defined as:

$$\mathcal{L} = \mathcal{K} - \mathcal{V} \tag{3.12}$$

where $\mathcal{T}$ is the kinetic energy and $\mathcal{V}$ is the potential energy. The Lagrangian is a function of the position $\mathbf{q}$ and momenta $\dot{\mathbf{q}}$ of all the particles ($\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}})$). The evolution of the system can be found by solving the Euler-Lagrange equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}}\right) - \frac{\partial \mathcal{L}}{\partial q} = 0 \tag{3.13}$$

In a system of particles with Cartesian coordinates $\mathbf{r_i}$, the kinetic energy has the form:

$$\mathcal{K} = \sum_{i=1}^{N} \frac{\dot{\mathbf{r}}_\mathbf{i}^{\,2} m_i}{2} \tag{3.14}$$

where m_i is the mass of particle i. The solution of Equation 3.13 becomes

$$m_i \ddot{\mathbf{r}}_\mathbf{i} = \mathbf{f}_\mathbf{i} \tag{3.15}$$

where $\mathbf{f}_\mathbf{i}$ is the total force on particle i and it applies to the center of mass of the particle. The force is calculated from the potential energy of particle i.

$$\mathbf{f}_\mathbf{i} = \frac{\partial}{\partial \mathbf{r}_\mathbf{i}} \mathcal{L} = -\frac{\partial}{\partial \mathbf{r}_\mathbf{i}} \mathcal{V} \tag{3.16}$$

Hamiltonian Mechanics is an alternative formulation to the Lagrangian Mechanics. The advantage of using Hamiltonian mechanics is that instead of solving the evolution of the system by a series of second order differential equations, it defines its dynamics in terms of first order equations, but twice as many of them. The Hamiltonian is defined by:

$$\mathcal{H} = \sum_k \dot{q}_k p_k - \mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}) \tag{3.17}$$

Since the potential is independent of the velocities and time, the Hamiltonian is equal to the total energy of the system:

$$\mathcal{H} = \mathcal{K} + \mathcal{V} \tag{3.18}$$

The difference with respect to the Lagrangian equations of motion is that the momenta features in the Hamiltonian form:

$$\dot{q}_k = \frac{\partial \mathcal{H}}{\partial p_k}$$
$$\dot{p}_k = -\frac{\partial \mathcal{H}}{\partial q_k} \tag{3.19}$$

In Cartesian coordinates, Hamilton's equations of motion become:

$$\dot{\mathbf{r}}_i = \frac{\dot{\mathbf{p}}_i}{m_i} \tag{3.20}$$
$$\dot{\mathbf{p}}_i = -\frac{\partial \mathcal{V}}{\partial \mathbf{r}_i} = \mathbf{f}_i \tag{3.21}$$

The solving of $6N$ first order differential equations is required in order to compute the trajectories of all the N particles in the system. Obtaining the exact trajectories for a typical simulation system of $N = 100 - 1000$ particles, however, is impossible. Simulations rely on numerically approximate solutions and tracing the evolution of the system proceeds in the following way. With the initial conditions for the system ($\mathbf{r}$ and $\mathbf{p}$ specified at $t = 0$), the forces on each particle are calculated via Equation 3.21. The instantaneous acceleration on each particle is then calculated by

$$\mathbf{a}_i = \frac{\mathbf{f}_i}{m_i} \tag{3.22}$$

where the force and thus acceleration are assumed to be constant over some short time interval, δt. The new positions and momenta are calculated, using the thus calculated accelerations in order to advance to a new point in phase space at $t = \delta t$. The step is

repeated this time, using the newly obtained positions and momenta. This 'step-by-step' advancement of the positions and momenta is performed using an algorithm that produces a trajectory through phase space that closely matches the actual (continuous) trajectory to a desired degree of accuracy. The algorithm for advancing the system in phase space can be tested on a system that is analytically solvable. Several schemes for integrating the equations of motion have been developed, including predictor-corrector algorithms, the Verlet algorithm, the leap frog and the velocity-Verlet, discussed in detail below. It is essential for the integration scheme to satisfy two important criteria - area preservation and time reversibility.

For isolated systems at equilibrium (where there is no explicit time dependence in the Hamiltonian), the Hamiltonian, which represents the total energy of the system, is a constant of motion.

$$\begin{aligned} \frac{d\mathcal{H}}{dt} &= \frac{\partial \mathcal{H}}{\partial q}\dot{q} + \frac{\partial \mathcal{H}}{\partial p}\dot{p} \\ &= \frac{\partial \mathcal{H}}{\partial q}\left(\frac{\partial \mathcal{H}}{\partial p}\right) + \frac{\partial \mathcal{H}}{\partial p}\left(-\frac{\partial \mathcal{H}}{\partial q}\right) \\ &= 0 \end{aligned} \tag{3.23}$$

The $\mathcal{H}$ remains constant and therefore the phase space characterizing the system must remain constant as well. That is why the integration scheme must be area preserving. It must not allow the total energy of the system to augment or diminish throughout the simulation. The trajectory in phase space must also be time reversible. If the momenta of all the particles in the system are reversed, the system should go back to its initial configuration following the exact same trajectory. If time reversibility condition is not

satisfied, that would mean that two initial state points can produce the same future state point in phase space, which is not compatible with the area preservation requirement.

The velocity-Verlet algorithm is chosen as an integrator of the equations of motion for its area preserving and time reversible properties. It advances the positions and velocities of the system's particles using both full time steps (for positions) and half time-steps (for velocities). The details are as follows:

Initially the position and velocities of the particles at time t are known. The accelerations are found using Equations 3.21 and 3.22.

1. The particles' positions at time t + δt are found by

$$\vec{r}_i(t+\delta t) = \vec{r}_i(t) + \vec{v}_i(\delta t)\delta t + \vec{a}_i\frac{1}{2}(\delta t) \tag{3.24}$$

The equation is a Taylor expansion around $\vec{r}$ and this first move is schematically depicted in Figure 3.7 (left).

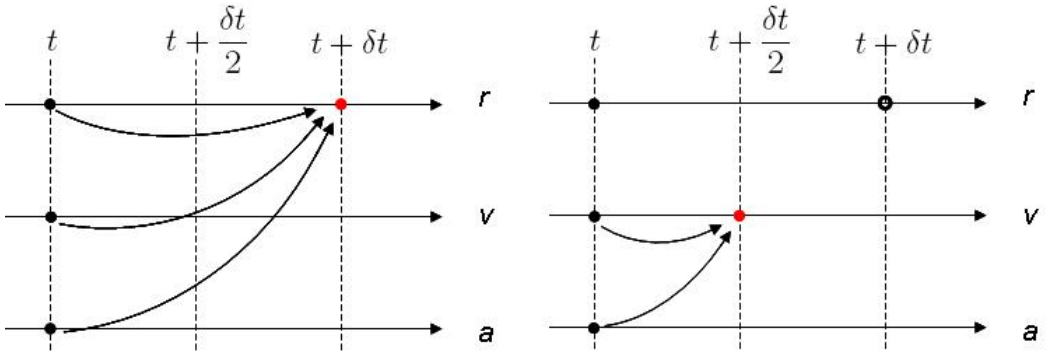


Figure 3.7. Schematic representation of the first two steps of the velocity-Verlet algorithm.

2. The velocities at half time-step are calculated

$$\vec{v}_i(t+\frac{\delta t}{2}) = \vec{v}_i(t) + \frac{\vec{a}_i\delta t}{2} \tag{3.25}$$

This essentially provides updated positions that are accurate to the degree that the velocities at the mid-point of the time step approximate the average velocities over the entire time step (Figure 3.7 right).

3. The forces and accelerations at $t + \delta t$ are once again computed via Equations 3.21 and 3.22 and the new positions (Figure 3.8 left).

4. and finally the velocities at $t + \delta t$ are found (Figure 3.8 right)

$$\vec{v}_i(t + \delta t) = \vec{v}_i(t + \frac{1}{2}\delta t) + \vec{a}_i(t + \delta t)\left(\frac{1}{2}\delta t\right). \tag{3.26}$$

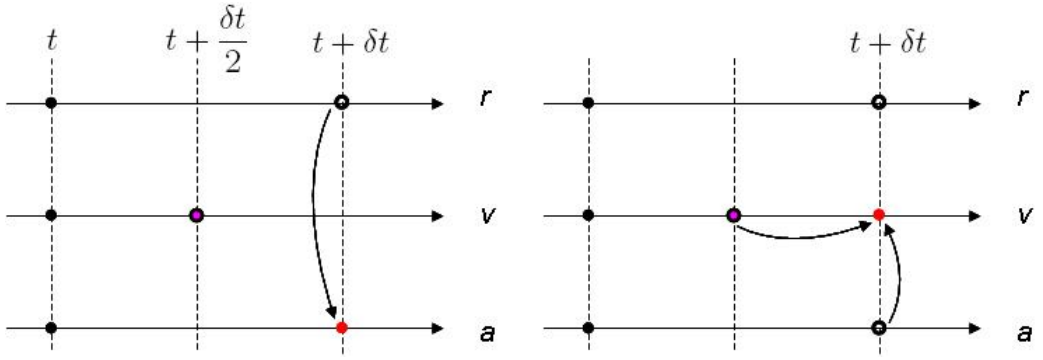


Figure 3.8. Schematic representation of the last two steps of the velocity-Verlet algorithm.

3.5.2 Hybrid Monte Carlo scheme

Conventional Monte Carlo sampling is performed only with a single particle update at each step. Random displacement of all the particles at once could result in low acceptance probability and consequently large number of steps before reaching equilibrium. At the same time, the size of each step cannot be taken to be too large, as such a move would again result in a low acceptance ratio. Therefore even though MC

sampling provides a complete exploration of phase space, the limitation of the step size means that the process will take too many iterations. This is particularly prominent for dense liquids. The step size in this case cannot be taken to be too large, as such a move will result in overlapping of the particles and consequently low acceptance probability. On the other hand MD simulations allow for larger time steps in phase space, because of the global updates of particles, subjected to Newton's equations of motion. Compared to MC however, MD sampling also suffers from limitations. MD moves are deterministic in nature, which means that unphysical moves cannot be performed. MD follows a direct slope of decreasing potential energy, which could lead to trapping the simulation on a limited cycle of phase space.[21] Furthermore, the time step in MD is limited by the energy conservation law, which is not the case in MC sampling.

In order to combine the advantages of the two methods, a hybrid technique consisting of both types of moves (MC and MD) has been designed by Duane and Kennedy, called Hybrid Monte Carlo (HMC).[63] Originally developed to simulate the behavior of quarks in the nucleon, the technique has been extensively used later for simulating polymers in a liquid state as well as dense liquid systems.[16,21,64,65]

The HMC algorithm is essentially stochastic in nature generating a Markov chain of states, thus resembling more the MC scheme. Nevertheless it provides global updating of the system with high acceptance probability and as such faster convergence. The global moves are performed in the configurational part of phase space, the difference with respect to MC being that the moves are deterministic - the next state in the Markov chain is generated using Hamilton's equations of motion. Because of the deterministic updating of the system, the outcome of the new configuration depends on the momenta of the particles

in the old configuration, and so the transition matrix, or conditional probability of having the new configuration by being in the old one has to be updated with a new term which is a function of the particles' momenta. For this purpose, it is convenient to use a Gaussian distribution, from which the initial momenta of the particles at the temperature of choice are drawn:

$$\alpha_{mn} \propto \alpha'_{mn} \exp\left(-\frac{1}{k_B T}\sum_{i=1}^{N}\frac{p_i^2}{2m}\right) \tag{3.27}$$

The MC acceptance criterion (Equation 5.20) then is changed to the following expression in the HMC scheme:

$$acc_{mn} = \min\left\{1, \frac{\rho_n \exp\left(-\frac{1}{k_B T}\frac{\mathbf{p_n^2}}{2m}\right)\alpha'_{nm}}{\rho_m \exp\left(-\frac{1}{k_B T}\frac{\mathbf{p_m^2}}{2m}\right)\alpha'_{mn}}\right\} = min\left\{1, \frac{\exp\left(\frac{-E_n(\mathbf{r},\mathbf{p})}{k_B T}\right)\alpha'_{nm}}{\exp\left(\frac{-E_m(\mathbf{r},\mathbf{p})}{k_B T}\right)\alpha'_{mn}}\right\} \tag{3.28}$$

Finally the expression can be simplified to

$$acc_{mn} = min\left\{1, \exp\left(\frac{\Delta E_{mn}}{k_B T}\right)\frac{\alpha'_{nm}}{\alpha'_{mn}}\right\} = min\left\{1, \exp\left(\frac{\Delta E_{mn}}{k_B T}\right)\right\} \tag{3.29}$$

where ΔE_{mn} is the change in total energy and $\mathbf{p_n}$ is the generalized momenta of the particles in state n. Since the transition is performed in a Markov chain, the microscopic reversibility condition has to be met, which means that $\alpha'_{mn} = \alpha'_{nm}$. As the transition from the old to the new configuration is performed by using a trajectory consisting of several MD steps, the above condition is met when the MD algorithm is time reversible - one which retraces its steps back to the old state when the signs of the momenta of the new state are reversed. This is because α_{nm}, which includes information about the momenta of

all particles, is invariant to the change in sign of the momenta. At the end of the MD trajectory, no matter if the move is accepted or not, a new Gaussian distribution for the momenta is generated. This is an essential requirement of the HMC algorithm, as the detailed balance condition has to be met. If this step is skipped, then the chance of returning to the old configuration is zero.

The initialization and the main HMC cycle can be schematically represented in Figure 3.9. Initialization starts with the choice of position and momenta (drawn from a Gaussian distribution) of the particles in the system. The total energy of the system is calculated (E_1) and a MD trajectory is performed. With the new position and momenta of the particles, the total energy is calculated once again (E_2) and the Boltzmann factor s is estimated ($s = \exp[(E_1 - E_2)/k_BT]$) and a random number ($r$) between 0 and 1 is generated. The new positions are accepted if $s > r$. Otherwise the old configuration is preserved. Regardless of the acceptance or rejection of the positions, new momenta are generated and the cycle is repeated. After performing a given number of cycles, the simulation ends.

It has to be noted that, even though the Hamilton's equations are used to trace the evolution of the system from the old to the new state, only the configurational phase space is explored. No real time dependent properties of the system can be traced with HMC, as the MD trajectory is used only to generate a new state and the probability distribution is governed by the detailed balance condition. As a result, the ensemble average does not depend on the time step chosen. This, however, is not the case for the acceptance probability. Even when the MD algorithm is time reversible and area preserving, too large time steps result in significant drifting of the total energy and therefore low acceptance

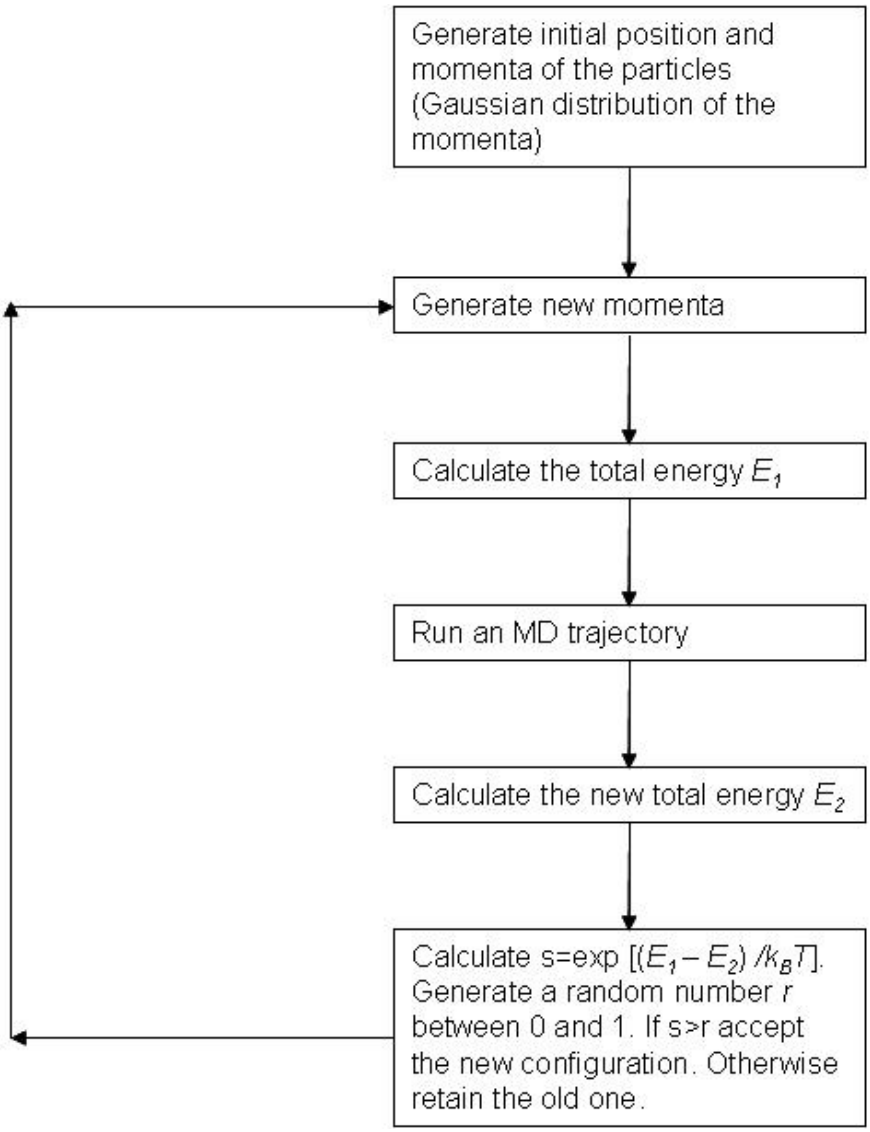


Figure 3.9. Schematic representation of the HMC algorithm.

probability (Equation 3.29). Nevertheless, compared to an MD scheme, there is no limitation imposed by the energy conservation law and larger time steps are usually chosen. Any area preserving and time reversible algorithm can be used instead of MD, the latter however is often the choice in HMC, as it provides reasonably high acceptance probability.[64]

The difference between the MC, MD and HMC methods in exploring phase space can be schematically presented as shown in Figure 3.10. The trajectory of a MC

simulation allows the sampling of the whole phase space; however, due to the local updating of the system, the process of exploring the phase space takes might take significant computational time. MD sampling is characterized with global updating and therefore big iteration steps. Some regions of phase space, however, may not be attempted, as sometimes the MD simulation can get stuck in small regions of phase space (it can be trapped in a cycle). HMC, however, manages with both issues presented, as global iterations are carried out to sample phase space faster than an MC simulation and random changes in the momenta of the particles do not allow the simulation to get trapped in certain regions of phase space. The HMC technique however cannot be used to sample efficiently states with very low probability, which play the role of bottlenecks in phase space. It can be viewed as a more efficient MC sampling technique, yet limited by the acceptance probability.

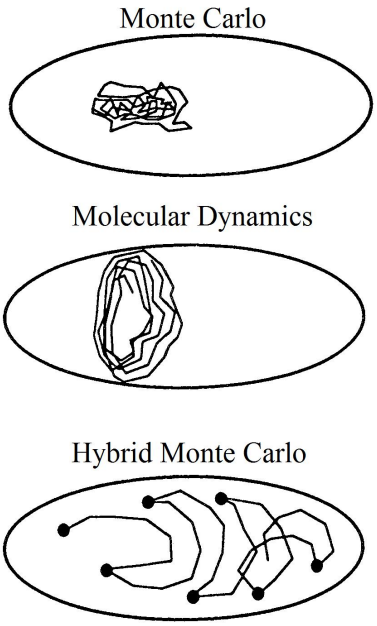


Figure 3.10. Schematic representation of the trajectories in phase space for an MC, MD and HMC sampling.[21]

CHAPTER IV

VAPOR-LIQUID EQUILIBRIA OF BRANCHED ALKANES

Phase behavior of hydrocarbons has been studied exhaustively, due to its importance in processes related to polymer, chemical and petrochemical industries. Information on their properties at coexistence is necessary for efficient extraction, design of the separation equipment and oil recovery processes. Therefore an exact knowledge of the phase diagrams of various alkane series is desired. Such phase diagrams are known well only for normal alkane series of up to 10 carbon atoms and for branched alkane series of up to 8 carbon atoms. Empirical correlations and semi-empirical equations of state can contribute to solving the issue. They however often rely on the critical properties of hydrocarbons, which are difficult to obtain experimentally, given the instability of the molecules. As a result, molecular simulations provide a good alternative for studying vapor-liquid equilibria of such fluids.

Vapor-liquid equilibria as well as critical properties of a large variety of hydrocarbons have been determined by the Gibbs ensemble Monte Carlo method. The method, however, relies on particle insertion and deletion moves between two simulation boxes in order to achieve the chemical potential equilibration. Therefore a biased insertion has to be implemented in order to facilitate the particle introduction in the dense liquid phase. The biasing strategy depends on the molecular architecture. In the case *n*-alkanes the configurational bias Monte Carlo (CBMC) technique is often used. This technique,

while well working for chain molecules, has been shown to be invalid for branched alkanes and a modification has been implemented in order to correct it.[66] Instead of growing multiple branches one after the other, the segments of the molecule are grown simultaneously at a branched point. This adds another dimension of complexity to the method and slows down the simulation as the acceptance probability of adding a branch to the linear chain decreases.

In our simulations of branched alkanes we use the HMC-WL method to construct a phase diagram of isobutane and isopentane. The insertion step of the GEMC method is avoided as we perform our simulation in the NPT ensemble. The HMC technique that is implemented in the WL scheme has been successfully applied in several cases in order to replace the translation, rotation and conformational changes moves within each of the simulation boxes of the GEMC simulation.[23,67,68]

4.1 Potential model

In our simulation study we use the NERD (Nath, Escobedo and de Pablo revised) force field developed for the simulation of straight and branched alkanes.[23,69] It has been parameterized for linear and branched alkanes and has shown to accurately reproduce experimental results. It is also a united atom potential model, for which a group of atoms is represented as a single interaction site, thus accounting for faster computation of the MD move.

Two types of interactions are considered in our system - bonding and non-bonding interactions. The non-bonding interactions are described by the classical Lennard-Jones 12-6 potential:

$$U(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{4.1}$$

The values of the parameters ε_i and σ_i are taken from previous estimations,[23] and applied to two different groups of atoms - CH_2 and CH_3. This potential is used for intermolecular interactions as well as interaction sites within the same molecule, separated by more than three bonds. Similarly to previous studies, the Lorentz-Berthelot combining rules are used for the parameters of unlike interaction sites.[70]

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_{ii} + \sigma_{jj}\right) \tag{4.2}$$

$$\varepsilon_{ij} = \left(\varepsilon_{ii}\ \varepsilon_{ij}\right)^{1/2} \tag{4.3}$$

The flexibility of the branched chain is represented by three different potentials accounting for the stretching, bending and torsional changes that occur in alkanes. The stretching potential has the following form:

$$V(r)/k_B = \frac{K_r}{2}(r - b_{eq})^2 \tag{4.4}$$

where K_r is a force constant, r is the distance between two neighboring sites and b_{eq} is the equilibrium distance between two neighboring sites (Figure 4.4). The bending potential is given by the following equation:

$$V(\theta)/k_B = \frac{K_\theta}{2}(\theta - \theta_{eq})^2 \tag{4.5}$$

where K_θ is a bending force constant, θ is the bending angle and θ_{eq} is the equilibrium angle (Figure 4.5). Finally, the torsional changes are modeled by the torsion potential:

$$V(\phi)/k_B = V_0 \; + \; V_1(1 + cos\ \phi) \; + \; V_2(1 - cos\ 2\phi) \; + \; V_3(1 + cos\ 3\phi) \qquad (4.6)$$

where the V_i $i = 1, 2, 3$ are dihedral constants and ϕ is the dihedral angle (Figure 4.6). The parameters for the bonded interaction potentials can be found in Table 4.1 and non-bonded Lennard-Jones parameters in Table 4.2.

Table 4.1. Bonding parameters for isobutane and isopentane

Bond stretching parameters			
K_r (KÅ^{-2})	b_{eq} (Å)		
96 500	1.54		
Bond bending parameters			
K_θ (K rad^{-2})	θ_{eq} (°)	θ_{eq} (°) (centered at a CH unit)	
62 500	114.0	109.4	
Torsional potential parameters			
V_0 (K)	V_1 (K)	V_2 (K)	V_3 (K)
1416.3	398.3	139.12	-901.2

Table 4.2. Lennard-Jones parameters for isobutane and isopentane

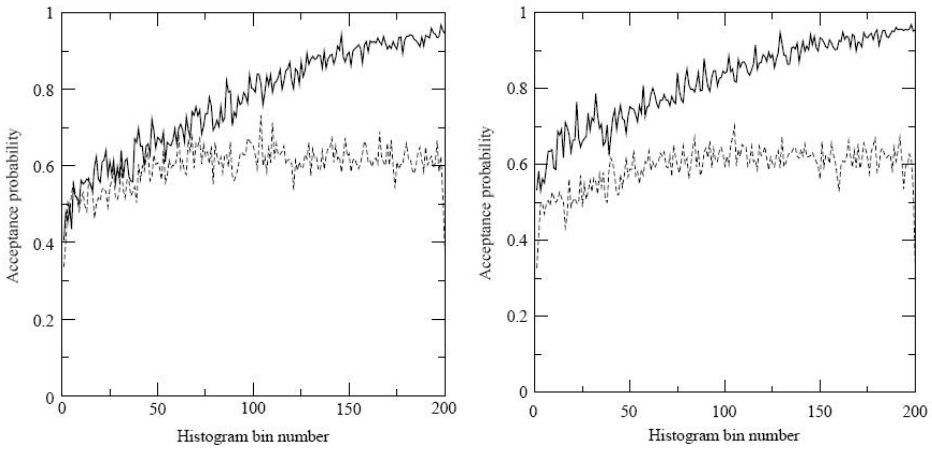
	σ_{CH_3} (Å)	σ_{CH_2} (Å)	σ_{CH} (Å)	ε_{CH_3} (K)	ε_{CH_2} (K)	ε_{CH} (K)
Isobutane	3.88	–	3.85	78.23	–	39.7
Isopentane	3.90	3.93	3.85	79.5	45.8	39.7

4.2 Parameter optimization

In our simulations we use a density range for isobutane and isopentane of 0.006 and 0.6 g/cm^3 and the corresponding volume range is divided into 200 bins. In accord with most of the MC simulations, the aim is to achieve a 50 % acceptance rate for the

random moves in our system.[70] A shortened simulation is performed for this purpose, with only one value of the convergence factor f. Since the total branched molecule is characterized with high frequency intramolecular vibrations, including the bond stretching, bond bending and conformational changes modes, and long range intermolecular interactions, the forces on a molecule are divided into a short distance and long distance forces ($\mathbf{F} = \mathbf{F}_{short} + \mathbf{F}_{long}$). The multiple time step RESPA-NVE algorithm, proposed by Martyna *et al.* is implemented in the HMC-WL scheme.[71] For the MD trajectories, the parameters that can be adjusted are the time step for the slow intermolecular and fast intramolecular motions, as well as the length of the MD trajectory (number of steps before applying the acceptance criterion[16]). In the case of isobutane, it is found that an MD length of 20 steps with a 19.3 fs time step for the slow intermolecular modes and fast intramolecular modes integrated 800 times over a single slow mode (800 times over 19.3 fs or time step of 19.3/800 = 0.024125 fs), gives an acceptance probability of 50 % or above. In the case of isopentane, the same criterion is fulfilled when the MD length is 20 steps with a 16.9 fs time step for the slow intermolecular modes and fast intramolecular modes integrated 700 times over a single slow mode (700 times over 16.9 fs or time step of 16.9/700 = 0.024143 fs).

The acceptance probabilities for both the MD trajectories and the random volume changes are traced as a function of the volume bin number. It was found that the lowest acceptance probability for both moves is for the smallest volume bin(Figure 4.1). This result can be anticipated for any random type of moves as, in this volume bin the particles' density is the highest and high energy state trials that are rejected can be easily generated.



(a) Isobutane at 300 K (b) Isopentane at 350 K

Figure 4.1. Acceptance probability for the two types of random moves. Solid line - MD trajectories; dashed lines - volume changes

4.3 Convergence criteria

In the WL method, the key point is to obtain an accurate estimate of the biasing function. In the NPT ensemble, it is the canonical partition function $Q(N, V, T)$ that is dynamically updated in order to approach its true values in the volume interval embracing the vapor and liquid phases. Therefore its convergence is traced during the simulation. As the convergence factor decreases after each WL-HMC sweep, $Q(N, V, T)$ becomes smoother. In Figure 4.2 the upper line corresponds to the running estimate of the partition function as a function of the reduced volume for the smallest convergence factor. The smoothness of the curve shows that an accurate estimate of $Q(N, V, T)$ is obtained.

In accord with the evolution of $ln(Q(N, V, T))$ with volume, the derivative of the logarithm of the running estimate of $Q(N, V, T)$ with volume also shows a smooth noise free curve at a low convergence factor (Figure 4.3). This can be viewed as another indicator of the simulation convergence.

Finally, the convergence can be quantified by tracing the relative error in

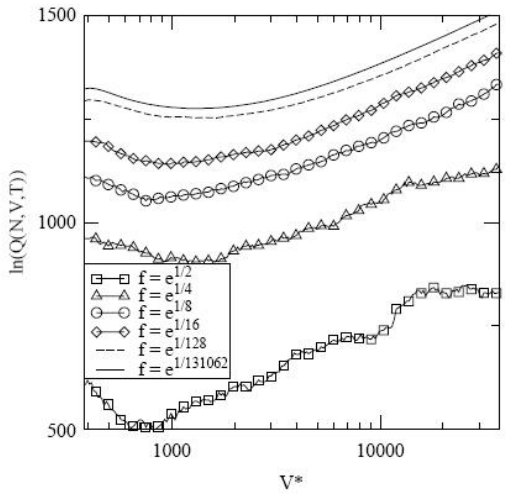


Figure 4.2. Evolution of the canonical partition function as a function of the reduced volume at different convergence factors.

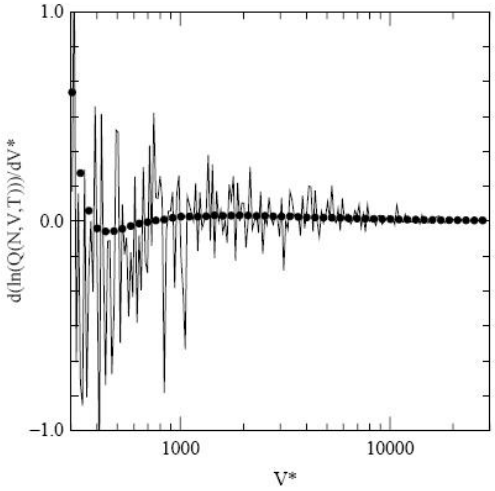


Figure 4.3. Derivative of the logarithm of the running estimate for the canonical partition function with respect to volume for isobutane at $T = 350$ K (solid line, $f = \sqrt{e}$; filled circles, $f = e^{1/2^{16})}$.

$ln(Q(N,V,T))$. It can be defined in the following expression:

$$\Delta E_n = \left| \frac{lnQ_{f_n}(N,V,T) - lnQ_{f_{n-1}}(N,V,T)}{lnQ_{f_n}(N,V,T)} \right| \tag{4.7}$$

where $Q_{f_n}(N,V,T)$ is the running estimate of the canonical partition function at convergence factor $f = e^{1/2^n}$. The convergence is once again confirmed in Figure 4.4 as

the relative error becomes insignificant at high n values ($n > 14$). At the end of the simulation, it becomes less than 10^{-4} for all values of the relative volume.

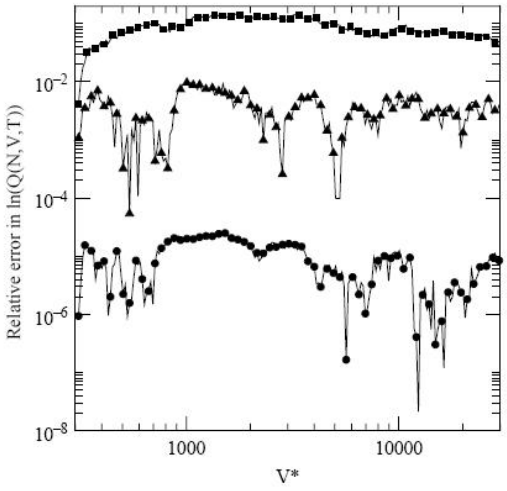


Figure 4.4. Relative error in the logarithm of the running estimate for the canonical partition function for isobutane at T = 350 K (ΔE_2, squares; ΔE_6, triangles; ΔE_{16}, circles)

From the evolution of the logarithm of the partition function as well as its derivative and relative error, it is clear that a low convergence factor ensures the convergence of the simulation with a final accurate estimate of $Q(N, V, T)$ as a function of volume.

4.4 Results and discussion

The WL-HMC simulations are carried out for eight temperatures ranging from $T = 290$ to $T = 360$ K for isobutane and the 340-410 range for isopentane. The vapor and liquid densities of isobutane and isopentane at coexistence are plotted in Figure 4.5 and Figure 4.6, respectively. As in the case of copper, the equilibrium densities are calculated using equations 3.7 and 3.8, using the accurate estimate of the canonical partition function. Data from experimental and simulations studies using the GEMC method

combined with the CBMC technique are plotted as well for comparison. All the data concerning the critical properties obtained from experimental studies and our simulation work are summarized in Table 4.3.

Table 4.3. Critical temperatures and densities of isobutane and isopentane.

Compound	$T_{c,sim}$ (K)	$T_{c,exp}$(K)[22]	$\rho_{c,sim}$ (kg/m^3)	$\rho_{c,exp}$(kg/m^3)[22]
Isobutane	404 ± 8	408.13	236 ± 15	221
Isopentane	457 ± 8	460.95	241 ± 15	234

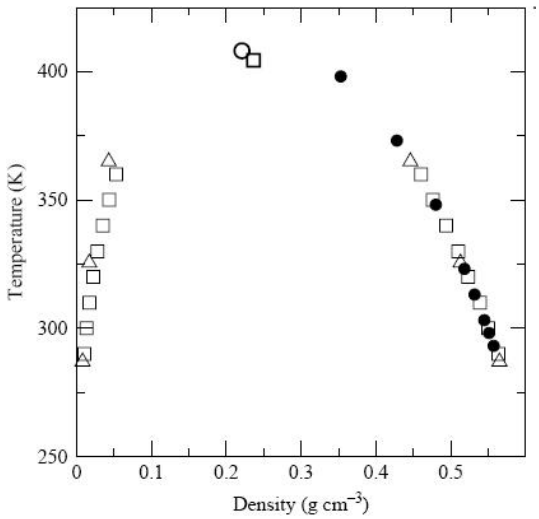


Figure 4.5. Vapour-liquid equilibria for isobutane obtained from NPT WL-HMC simulations (open squares), experiments[22] (filled circles) and Nath and de Pablo[23] simulations (open triangles). The bold symbols represent the critical point.

The WL-HMC simulation results are in excellent agreement with the available experimental data[22] and previous simulations.[23] Critical properties from our simulation are also compared to the experimental ones. For isobutane we found T_c = 404 ± 8 K, while ρ_c = 236 ± 15 kg/m^3. This values are within 1 and 7 % respectively of their experimental counterparts. The critical point of isopentane is found at T_c = 457 ± 8 and ρ_c = 241 ± 15. These values are even closer to experimental ones and differ only by 1 and

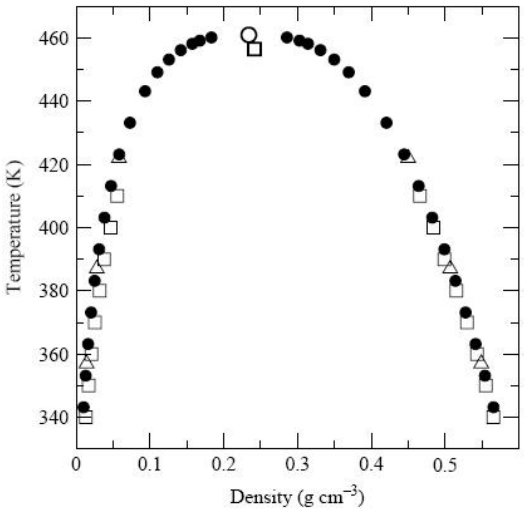


Figure 4.6. Vapour-liquid equilibria for isopentane obtained from NPT WL-HMC simulations (open squares), experiments[22] (filled circles) and Nath and de Pablo[23] simulations (open triangles). The bold symbols represent the critical point.

3 % for the critical temperature and density respectively. The densities for the two coexisting phases, obtained from the WL-HMC simulations, as well as the critical points for isobutane and isopentane, extrapolated from the WL-HMC simulations, are therefore in excellent agreement with the experimental data.

4.5 Multiple time step algorithm

In molecular dynamics simulations, the time step has to be chosen in accordance with the frequency of the intramolecular vibrations of the molecule. The adequate sampling of phase space is observed only when the time step is shorter than the period of the fastest vibrations. The problem that arises is that MD simulations of molecular fluids become too long compared to that of atomic fluids. The solution is to integrate the forces related to the high frequency intramolecular vibrations with shorter time steps and the intermolecular forces with a longer time step - multiple time step integration.

The RESPA-NVE multiple time step method proposed by Martyna *et al.* is used in

our simulations.[71] It is based on the Trotter factorization of the Liouville operator. The latter is defined as follows

$$iL = \frac{\partial \mathcal{H}}{\partial \mathbf{p}} \frac{\partial}{\partial \mathbf{r}} - \frac{\partial \mathcal{H}}{\partial \mathbf{r}} \frac{\partial}{\partial \mathbf{p}} = \{..., \mathcal{H}\} \tag{4.8}$$

where $\{..., \mathcal{H}\}$ is the Poisson bracket. Each configuration (point) in phase space is characterized with a multidimensional vector $\mathbf{\Gamma}(\mathbf{r}, \mathbf{p})$, which is depends on the position and momenta of all the particles in the system. In order to trace the evolution of $\mathbf{\Gamma}$ in phase space, we can write

$$\frac{\mathrm{d}\mathbf{\Gamma}}{\mathrm{d}t} = \{\mathbf{\Gamma}, \mathcal{H}\} \tag{4.9}$$

which gives the Hamilton's differential equations of motion (Equations 3.20 and 3.21), since for a single particle we have

$$\begin{aligned} \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} &= \{\mathbf{r}, \mathcal{H}\} \\ &= \frac{\partial \mathbf{r}}{\partial \mathbf{r}} \frac{\mathbf{p}}{m} + \frac{\partial \mathbf{r}}{\partial \mathbf{p}} \mathbf{F}(\mathbf{r}) \\ &= \frac{\mathbf{p}}{m} \end{aligned} \tag{4.10}$$

and

$$\begin{aligned} \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} &= \{\mathbf{p}, \mathcal{H}\} \\ &= \frac{\partial \mathbf{p}}{\partial \mathbf{r}} \frac{\mathbf{p}}{m} - \frac{\partial \mathbf{p}}{\partial \mathbf{p}} \frac{\mathrm{d}U}{\mathrm{d}\mathbf{r}} \end{aligned}$$

$$= \mathbf{F}(\mathbf{r}) \tag{4.11}$$

Therefore, the equations of motion can be written in the Liouville operator mode

$$\frac{\mathrm{d}\mathbf{\Gamma}}{\mathrm{d}t} = iL\mathbf{\Gamma} \tag{4.12}$$

The equations of motion can be solved to yield $\mathbf{\Gamma}$ at time t

$$\mathbf{\Gamma}(t) = \exp(iLt)\mathbf{\Gamma}(0) \tag{4.13}$$

Since the evolution of the system cannot be traced analytically, an approximation of the true operator accurate at $\Delta t = t/p$ is applied p times

$$\mathbf{\Gamma}(t) = \prod_{i=1}^{p} \left(\exp(iL\Delta t)\right) \mathbf{\Gamma}(0) \tag{4.14}$$

In the NVE ensemble the Liouville operator can be separated into two parts

$$\begin{aligned} iL &= iL_{\mathbf{r}} + iL_{\mathbf{p}} \\ &= \frac{\mathbf{p}}{m}\frac{\partial}{\partial \mathbf{r}} + \mathbf{F}(\mathbf{r})\frac{\partial}{\partial \mathbf{p}} \end{aligned} \tag{4.15}$$

The step by step integration of motion can be solved by applying the Trotter formula.[72]

$$\exp\left(iL\Delta t\right) \approx \exp\left(iL_{\mathbf{p}}\Delta t/2\right) \exp\left(iL_{\mathbf{r}}\Delta t\right) \exp\left(iL_{\mathbf{p}}\Delta t/2\right) \tag{4.16}$$

The positions and momenta at time Δt are found by applying the Liouville operator in the

above written form under the initial conditions $(\mathbf{r}(0), \mathbf{p}(0))$. With the factorized form of the operator, position and momenta at time Δt are obtained after three steps:

$$\begin{aligned}
\exp\left(iL\Delta t\right)\left(\mathbf{\Gamma}\left[\mathbf{r}(0),\mathbf{p}(0)\right]\right) &= \exp\left(iL_{\mathbf{p}}\Delta t/2\right)\exp\left(iL_{\mathbf{r}}\Delta t\right)\exp\left(iL_{\mathbf{p}}\Delta t/2\right)\mathbf{\Gamma}\left[\mathbf{r}(0),\mathbf{p}(0)\right] \\
&= \exp\left(iL_{\mathbf{p}}\Delta t/2\right)\exp\left(iL_{\mathbf{r}}\Delta t\right)\mathbf{\Gamma}\left[\mathbf{r}(0),\mathbf{p}(0)+\mathbf{F}(0)\frac{\Delta t}{2}\right] \\
&= \exp\left(iL_{\mathbf{p}}\Delta t/2\right)\mathbf{\Gamma}\left[\mathbf{r}(0)+\frac{\mathbf{p}(\Delta t/2)}{m}\Delta t,\mathbf{p}(0)+\mathbf{F}(0)\frac{\Delta t}{2}\right] \\
&= \mathbf{\Gamma}\left[\mathbf{r}(0)+\frac{\mathbf{p}(\Delta t/2)}{m}\Delta t,\mathbf{p}(0)+\mathbf{F}(0)\frac{\Delta t}{2}+\mathbf{F}(\Delta t)\frac{\Delta t}{2}\right] \quad (4.17)
\end{aligned}$$

Thus the evolution in the position and momenta is followed by the velocity-Verlet algorithm (Figure 3.7 and 3.8), the only difference being that the order of the first two steps with respect to the previous description of the algorithm is inversed. The factorized operators are applied sequentially by updating first the momenta of the particles, with step $\Delta t/2$, then the positions with step Δt and finally again the momenta with a step $\Delta t/2$. In a similar way, the Liouville operator $iL_{\mathbf{p}}$ can be broken into two parts when considering the forces due to short range interactions ($\mathbf{F}_{short}$) and those due to long range interactions ($\mathbf{F}_{long}$). In terms of velocities, the first part is written as

$$iL_{short} = \frac{\mathbf{F}_{short}}{m}\frac{\partial}{\partial \mathbf{v}} \quad (4.18)$$

where the the force $\mathbf{F}_{short}$ is due to short range intramolecular interactions, including the bond stretching, bond bending and torsional changes. The second part of the operator can

be written as

$$iL_{long} = \frac{\mathbf{F} - \mathbf{F}_{short}}{m}\frac{\partial}{\partial \mathbf{v}} = \frac{\mathbf{F}_{long}}{m}\frac{\partial}{\partial \mathbf{v}}. \tag{4.19}$$

The total Liouville operator thus becomes

$$iL = iL_{\mathbf{r}} + iL_{short} + iL_{long} \tag{4.20}$$

The separation of the time scales is obtained by introducing a reference system characterized with the fast motions due to short range particle interactions with a Liouville operator $iL_{ref} = iL_{\mathbf{r}} + iL_{short}$. The time step for this system is chosen in accord with the fastest vibrations by dividing the large time step suitable for long range interactions into n $\delta t = \Delta T/n$. The remaining term of the total Liouville operator iL_{long} advances the velocities over time Δt. The Trotter expansion reads

$$\begin{aligned}
\exp(iL\Delta t) &= \exp\left[\left(iL_{\mathbf{r}} + iL_{short} + iL_{long}\right)\Delta t\right] \\
&= \exp\left(iL_{long}\frac{\Delta t}{2}\right)\exp\left[\left(iL_{short} + iL_{\mathbf{r}}\right)\Delta t\right]\exp\left(iL_{long}\frac{\Delta t}{2}\right) \\
&= \exp\left(iL_{long}\frac{\Delta t}{2}\right)\left[\exp\left(iL_{short}\frac{\delta t}{2}\right)\exp\left(iL_{\mathbf{r}}\delta t\right)\exp\left(iL_{short}\frac{\delta t}{2}\right)\right]^{n}\exp\left(iL_{long}\frac{\Delta t}{2}\right)
\end{aligned}$$

The integration of the equations of motion proceeds in the following way

1. First the velocities are advanced over time step $\Delta t/2$ due to the long range force $\mathbf{F}_{long}$.

2. The equations of motions are integrated with the velocity-Verlet algorithm n times over time step δt under the influence of the short range force $\mathbf{F}_{short}$.

3. The velocities are once again advanced over time step $\Delta t/2$ due to the long range force $\mathbf{F}_{long}$, this time a s a function of the new positions obtained in the previous step.

CHAPTER V

VAPOR-LIQUID EQUILIBRIA OF LONG CHAIN NORMAL ALKANES

5.1 Potential model

The simulations of long chain alkanes are based on the TraPPE united atom model with paramters introduced by Martin and Siepmann.[33] Similarly to the NERD force field, the force field consists of bonding and non-bonding interactions. The latter are modeled after the Lennard-Jones potential (Equation 4.1) and are used to describe interactions between interaction sites of two different molecules as well as interactions between sites on the same molecule, separated by at least three bonds.

The bonding interactions are modeled using a bond-bending potential

$$V(\theta)/k_B = \frac{K_\theta}{2}(\theta - \theta_{eq})^2 \tag{5.1}$$

where K_θ is a bending force constant, θ is the bending angle and θ_{eq} is the equilibrium angle and a torsional potential with the following form:

$$V(\phi)/k_B = V_1(1 + cos\ \phi)\ +\ V_2(1 - cos\ 2\phi)\ +\ V_3(1 + cos\ 3\phi) \tag{5.2}$$

where the V_i $i = 1, 2, 3$ are dihedral constants and ϕ is the dihedral angle. Parameters for the three potential models are taken from prior work[33] and presented in Table 5.1.

Table 5.1. Bonding and non-bonding parameters for icosane, tetracosane and triacontane

Parameters for the Lennard-Jones potential			
σ_{CH_3} (Å)	σ_{CH_2} (Å)	ε_{CH_3} (K)	ε_{CH_2} (K)
3.75	3.95	98	46
Bond bending parameters			
K_θ (K rad^{-2})	θ_{eq} (°)		
62 500	114.0		
Torsional parameters			
V_1 (K)	V_2 (K)	V_3 (K)	
355.03	-68.19	791.32	

5.2 Wang-Landau configurational bias Monte Carlo simulations

The long chain *n*-alkanes are simulated integrating the configurational bias Monte Carlo technique into the Wang-Landau scheme (WL-CBMC) in the isothermal-isobaric ensemble. Apart the volume changes that are sampled according to the biased probability (Equation 2.62) there are three more random moves related to exploring phase space. A molecule is picked at random and the following moves are performed

- translation (32.8 % of the attempted moves)
- rotation (32.8 % of the attempted moves)
- conformational change (32.8 % of the attempted moves)
- volume change (1.6 % of the attempted moves)

The probability of accepting any of the first two moves is calculated using the Boltzmann weight factor:

$$\text{acc}(o \rightarrow n) = min\left[1, \exp(-(U_n - U_o)/k_B T)\right] \tag{5.3}$$

In order to sample efficiently the conformational changes of the molecule, the configurational bias Monte Carlo (CBMC) scheme is used. It consists of regrowing part of the molecule in such a way that energetically favorable regions are explored during the regrowth. The Rosenbluth factor W is calculated for the new and the old conformation and the acceptance criterion is given by

$$\mathrm{acc}(o \rightarrow n) = min\left[1, \frac{W(n)}{W(o)}\right]. \tag{5.4}$$

Our simulation system consists of 100 molecules for n-icosane , n-tetracosane and n-triacontane. The minimum and maximum limits of the volume for the three systems are set at 0.0001 and 0.7 g/cm^3. The volume range is divided into 400 bins in uniform intervals over $\ln V$. Same methodology is followed as in section 2.3.2. In line with previous simulations, the convergence factor is chosen to be such that $\ln f = e$ at the beginning of the simulation and $\ln f = 10^{-5}$ for the last sweep.

5.3 Results and discussions

The coexistence densities at the specified temperature for the three n-alkanes used are obtained first by estimating the isothermal-isobaric partition function using Equation 3.5. It is used, in turn, to find the volume probability distribution given by

$$p(V) = \frac{Q(N, V, T)\exp(-PV/k_BT)}{Q(N, P, T)} \tag{5.5}$$

The coexisting pressure (P_{coex}) is found by assuring that the areas of the two peaks corresponding to the vapor and liquid phases are the same

$$\int_0^{V_b} p(V)\mathrm{d}V = \int_{V_b}^{\infty} p(V)\mathrm{d}V \tag{5.6}$$

where V_b correspond to the volume bin with minimum probability density. We plot in Figure 5.1 the volume distribution for *n*-icosane at T = 603 K at coexistence when the the above requirement is fulfilled and the same area is found under each peak.

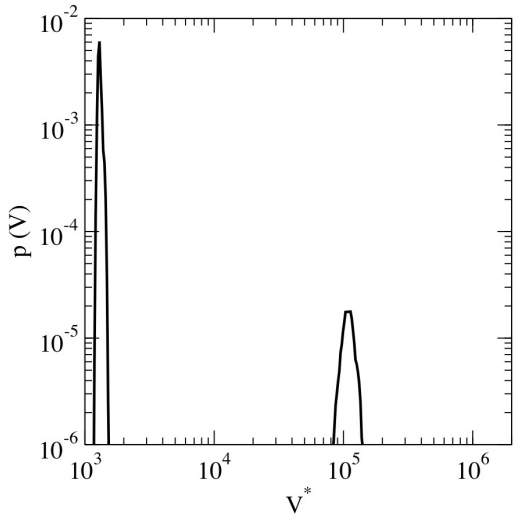


Figure 5.1. Volume distribution at coexistence for *n*-icosane at T=603 K. V^* denotes that the volume is reduced with respect to the Lennard-Jones parameters of the CH_2 group.

The simulation data of equilibrium pressure and temperature is presented in Figure 5.2. Our results are in good agreement with experimental data, slightly shifted to the right side of the experimental curve. The standard boiling point is determined by fitting the results of equilibrium pressure and temperature in Antoine's law (Equation 3.11). The results from our simulations show a boiling point of T_b = 595 ± 5 K for *n*-icosane which slightly underestimates the experimentally found boiling point of T_b^{exp} = 616.2 K.[73] The boiling point of *n*-tetracosane was found to be T_b = 647 ± 5 K, which is also lower but close to the experimental value of T_b^{exp} = 664 K.[74] In the case of *n*-triacontane, a value of

T_b = 710 ± 6 K is obtained from our simulations, which is in a good agreement with the experimental one of T_b^{exp} = 723 K.[74] In all three cases our data underestimate the experimental results of the boiling points, which reaches 3 % for *n*-icosane and *n*-tetracosane and 2 % for *n*-triacontane. The results are consistent with the GEMC results for shorter *n*-alkanes with the same potential model, which also show a deviation of 2 to 3 % (2% for *n*-dodecane and 3 % for *n*-octane).[33] The fact that WL-CBMC method showed the same order of deviation for *n*-alkanes, which are much longer than the previously simulated by Martin and Siepmann,[33] proves the reliability of the method. The difficulties associated with particle insertion and deletion moves are eliminated and reliable results can be obtained for long chains.

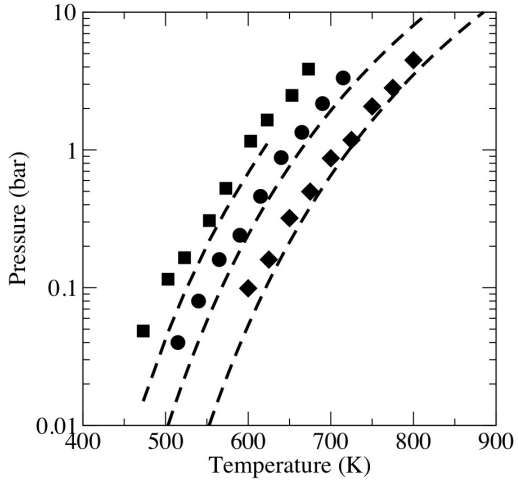


Figure 5.2. Saturation pressure as a function of temperature for *n*-icosane (filled squares), *n*-tetracosane (filled circles) and *n*-triacontane (filled diamonds). Our simulation data is shown as symbols, while experimental data as dashed lines.

5.4 Technical details

5.4.1 Monte Carlo sampling

The original Monte Carlo method is based on selecting random numbers from a uniform distribution. The MC integration method can be best illustrated by the ”hit and miss” technique for evaluating the numerical value of π. This can be achieved by estimating the area of a circle inscribed in a square (Figure 5.3). Both figures share the same origin and the radius of the circle r is half the side of the the square a. Pairs of random numbers (ξ with $0 < \xi < a$) are generated according to a uniform distribution, so that each pair corresponds to a point inside the square. Since the pairs are generated randomly, the picture in Figure 5.3, resembles a target with shots evenly distributed inside the area of the square, with a fraction of them being outside the circle.

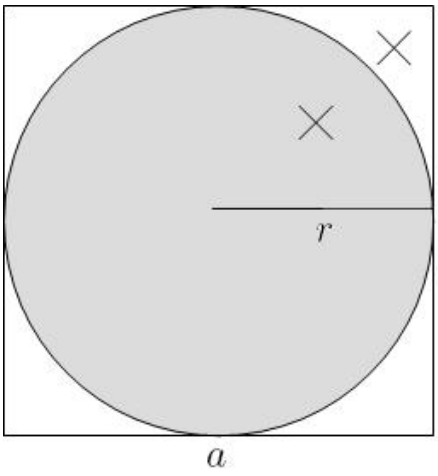


Figure 5.3. Hit and miss integration method.

Each time a point is generated, the distance between the center and the point is calculated. If the distance is shorter than the radius of the of the circle, a hit is recorded. The total number of hits τ_{hits} and shots (total number of generated pairs or τ_{shots}) are noted down. The area of the circle is then proportional to τ_{hits} and that of the square to τ_{shots}.

The ratio of the two areas can be given as:

$$\frac{S_{circle}}{S_{square}} = \frac{\pi r^2}{a^2} \approx \frac{\tau_{hit}}{\tau_{shot}} \tag{5.7}$$

Since $a = 2r$, an estimate of π can be calculated according to:

$$\pi \approx \frac{4\tau_{hit}}{\tau_{shot}} \tag{5.8}$$

The accuracy of the estimate depends on the number of shots, generated from the uniform distribution. The π number generated with 10^7 shots is found to be correct to four significant figures and to have one more decimal place would require an order of magnitude increase in the number of shots.[70]

Sampling the phase space requires a more efficient method of generating states. This could be achieved through generating configurations according to a uneven distribution, each configuration being chosen with its appropriate weight factor, known as importance sampling. The efficiency of the importance sampling can be illustrated by an example of evaluation of one dimensional integral.[2] Let us assume that the integral has the form:

$$I = \int_0^1 \mathrm{d}x \; f(x) \tag{5.9}$$

The idea is to estimate the integral by MC sampling, but instead of generating points according a uniform distribution, the points are distributed according to a nonnegative weight $w(x)$ in the interval [0,1]. Equation 5.9 can be rewritten as:

$$I = \int_0^1 \mathrm{d}x\, w(x)\, \frac{f(x)}{w(x)} \tag{5.10}$$

If $w(x)$ is the derivative of another function ($\mathrm{d}u(x) = \mathrm{d}x\, w(x)$), which is nonnegative and nondecreasing with $u(0) = 0$ and $u(1) = 1$, the above expression can be rewritten as:

$$I = \int_0^1 \mathrm{d}u\, \frac{f\,[x(u)]}{w\,[x(u)]} \tag{5.11}$$

The integral can be evaluated by generating K random values of u, uniformly distributed in the interval [0, 1]. The integral estimate is:

$$I \approx \frac{1}{K} \sum_{i=1}^{K} \frac{f\,[x(u_i)]}{w\,[x(u_i)]} \tag{5.12}$$

The weight $w\,[x(u_i)]$ plays an important role in the integral evaluation. This can be seen by calculating the variance of the integral ($Var(I)$). In Equation 5.12, the integral is represented as a sum of uncorrelated random variables. Therefore the variance of the sum can be expressed as the sum of the variances of each term:

$$Var(I) = Var\left(\frac{1}{K} \sum_{i=1}^{K} \frac{f\,[x(u_i)]}{w\,[x(u_i)]}\right) = \frac{1}{K^2} \sum_{i=1}^{K} Var\left(\frac{f\,[x(u_i)]}{w\,[x(u_i)]}\right) \tag{5.13}$$

Using the definition of variance, for each term we can write:

$$\begin{aligned} Var(I) = \frac{1}{K^2} \sum_{i=1}^{K} \left\langle \left(\frac{f\,[x(u_i)]}{w\,[x(u_i)]} - \left\langle \frac{f}{w} \right\rangle \right)^2 \right\rangle \\ = \frac{1}{K} \left(\left\langle \left(\frac{f}{w} \right)^2 \right\rangle - \left\langle \frac{f}{w} \right\rangle^2 \right) \end{aligned} \tag{5.14}$$

From the latter expression, it can be seen that in the ideal case $f(x)/w(x)$ is a constant over the integration interval with a variance equal to zero. If on the other hand $w(x)$ is constant, then it would be a simple uniform Monte Carlo on the [0, 1] interval. In this case, if only a small fraction of f contributes significantly to the integral value (for example when $f(X)$ is very narrow and steep bell like curve), the variance would be proportional to $1/(Kf)$ and consequently result in a highly inaccurate estimation. When sampling the configurational phase space, most of the randomly constructed states will have very high energy as the number of configurations with particular energy increases exponentially with energy. Thus if using the brute force MC, sampling would be ineffective since a very large number of trials would have to be generated to have a good estimate of the average macroscopic quantity. The importance sampling of the configurational phase space has to be carried out with the weight function being proportional to the Boltzmann factor. The issue presented is that a transformation such as that from Equation 5.10 to Equation 5.11 is a difficult task and a necessary condition would be to calculate the partition function of the system under study.

It is possible to generate a sequence of random states so that at the end each state has occurred with its appropriate probability, without calculating the partition function of the system. This is achieved by setting up a chain of states, called a Markov chain, with a limiting distribution of $\rho_{N,V,T}$. In a Markov chain the outcome of each stochastic trial belongs to a finite set of trials, which is the configurational space that is explored $\{\Gamma_1,\ \Gamma_2,\ \Gamma_3, ... \Gamma_m,\ \Gamma_n....\}$ and the outcome of each trial depends only on the preceding one.[70] The probability of going from state m to a new one n is given by the transition probability π_{mn}, also known as the transition matrix. In a Markov chain, any process will

converge to a unique state with a limiting distribution:

$$\sum_m \rho_m \pi_{mn} = \rho_n \tag{5.15}$$

The Markov chain is constructed such that every state can be visited from another state, which is consistent with the ergodic hypothesis. The transition probability can be considered as the product of the probability to perform the move (transition matrix, denoted α_{mn}) and the probability of accepting the move (acceptance probability, denoted acc_{mn}). For convenience, the equation of the limiting distribution can be replaced with the unnecessary strong condition of microscopic reversibility, known also as detailed balance condition:

$$\rho_m \pi_{mn} = \rho_n \pi_{nm} \tag{5.16}$$

or

$$\rho_m \alpha_{mn} acc_{mn} = \rho_n \alpha_{nm} acc_{nm} \tag{5.17}$$

The transition matrix α_{mn} in the above equations is stochastic, consistent with the ergodicity requirement. It is designed so that the system can go from sate m to any of the neighboring states n with equal probability ($\alpha_{mn} = \alpha_{nm}$).

The solution of the microscopic reversibility equation as suggested by Metropolis is divided into two distinct cases:

1.

$$\text{if } \rho_n \geq \rho_m \ \rightarrow \ \pi_{mn} = \alpha_{mn} \text{ for } m \neq n \tag{5.18}$$

2.

$$\text{if } \rho_n < \rho_m \;\rightarrow\; \pi_{mn} = \alpha_{mn}(\rho_n/\rho_m) \text{ for } m \neq n \tag{5.19}$$

Thus in the first case, when $\rho_n \geq \rho_m$, $acc_{mn} = 1$ and in the second case when $\rho_n < \rho_m$, $acc_{mn} = \rho_n/\rho_m$. The two cases can be summarized in the following equation:

$$acc_{mn} = \min\left\{1, \frac{\rho_n}{\rho_m}\right\} \tag{5.20}$$

The above equation is a simplified form of the Metropolis acceptance criterion, due to the symmetry of the transition matrix. A more general form would include α_{mn}:

$$acc_{mn} = \min\left\{1, \frac{\rho_n \alpha_{nm}}{\rho_m \alpha_{mn}}\right\} \tag{5.21}$$

The first step in the MC scheme is to choose randomly a particle i in a simulation box of N particles. A neighboring state can be chosen at random within a certain distance from particle i. The translation of particle i from state m to state n is illustrated in Figure 5.4. The new position belongs to a finite number of possible positions N_p within the displacement volume, shown in Figure 5.4 as a shaded area. Thus, the transition matrix can be defined as:

$$\alpha_{mn} = 1/N_p \tag{5.22}$$

The transition probability of moving from state m to state n is then given according to Equations 5.18 and 5.19. When $\rho_n \geq \rho_m$, Equation 5.18 applies. The inequality is satisfied when $\delta U_{mn} = U_n - U_m \leq 0$ and the move is accepted. The calculation of the

potential energy difference does not require the complete calculation of each pair of particles, but just the changes associated with the translated atom:

$$\delta U_{mn} = \left(\sum_{j=1}^{N} U(r_{ij}^{n}) \; - \; \sum_{j=1}^{N} U(r_{ij}^{m}) \right) \tag{5.23}$$

If the move results in $\rho_n > \rho_m$, then Equation 5.19 applies. The attempted move is accepted this time with probability ρ_n/ρ_m. In calculating the ratio, it can be noted that the configurational integral Z does not figure in the acceptance criteria:

$$\frac{\rho_n}{\rho_m} = \frac{exp\,(-U_n/k_BT)}{Z_{N,V,T}} \frac{Z_{N,V,T}}{exp\,(-U_m/k_BT)} = \exp\,(-\delta U_{nm}/k_BT) \tag{5.24}$$

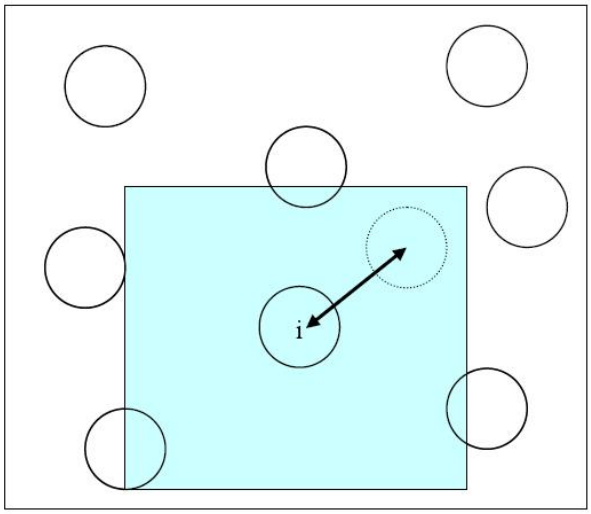


Figure 5.4. State m is obtained from state n by moving the particle within the shaded area (a two dimension representation of the displacement volume)

In order to accept or reject the move with transition probability $\exp\,(-\delta U_{nm}/k_BT)$, a random number ξ between 0 and 1 is generated and compared to the acceptance expression. If $\xi < \exp\,(-\delta U_{nm}/k_BT)$, then the move is accepted and if $\xi > \exp\,(-\delta U_{nm}/k_BT)$ - rejected. In the second case scenario of rejection of the move, the

old position of particle i is retained and the system remains in state m. Nevertheless it is important to count this state once again as part of the Markov chain. The transition probability of a Markov chain is such that:

$$\sum_{n} \pi_{mn} = 1 \tag{5.25}$$

This can be considered as a normalization requirement.[2] The probability of accepting the old configuration is then:

$$\pi_{mm} = 1 - \sum_{n \neq m} \pi_{mn} \tag{5.26}$$

which implies that the old configuration should be counted. Schematically the steps of the algorithm are shown in Figure 5.5.

Finally any thermodynamic property $\langle A \rangle_{N,V,T}$ is obtained by averaging over the n trials in the Markov chain:

$$\langle A \rangle_{N,V,T} = \langle A \rangle_{run} \tag{5.27}$$

Some important details of the Monte Carlo scheme include the displacement volume and the acceptance ratio of accepted moves. If the displacement volume is chosen to be very small, the particles are allowed to move over small distances and it will take many MC steps to arrive at a statistically independent configuration. On the other hand, if the volume is chosen to be too large, the particle position on average will change too much. This will lead to a strong increase in the energy of the system and most of the time the new configurations will be rejected. The displacement volume should be adjusted in

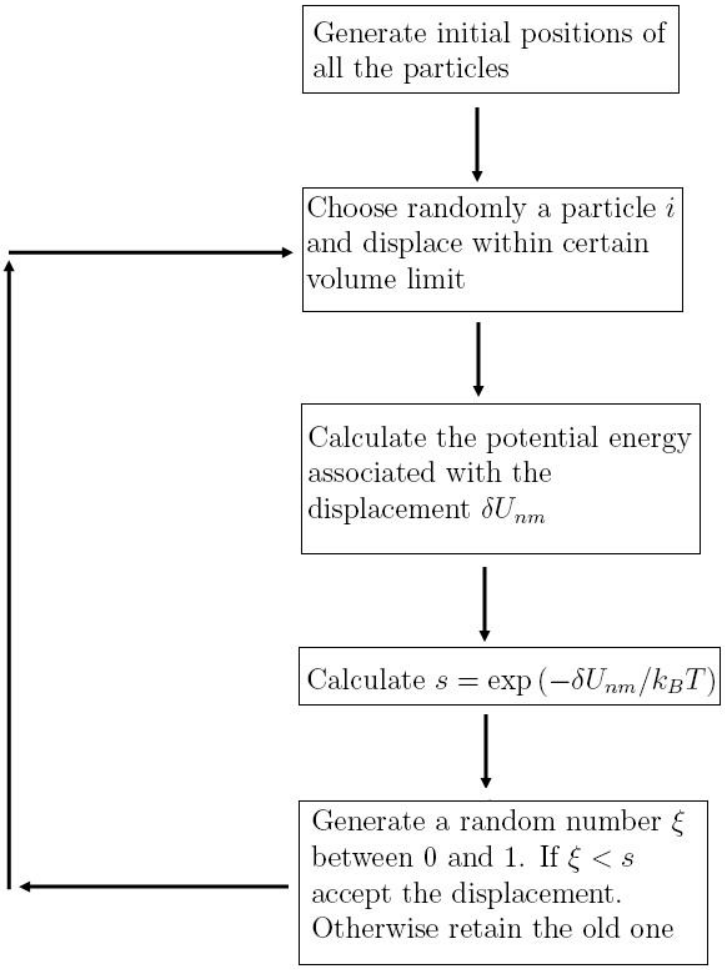


Figure 5.5. Schematic representation of the steps in the Metropolis Monte Carlo algorithm

accordance with the potential energy between the particles, such that a reasonable acceptance ratio is obtained. In MC simulations it is generally accepted a ratio of 0.5.[70]

An important requirement in devising the MC scheme is the detailed balance condition (Equation 5.16). Even though it is not strictly required for a Markov chain, its nonobservance often leads to wrong sampling. An example of such breaking of the detailed balance condition is provided in Frenkel's book on molecular simulations.[2] In the step consisting of the choice for the new configuration if the random displacement is going only in the positive direction (the randomly chosen x, y and z displacement vectors are

positive), then the reserve move of putting the particle back at its original configuration is impossible ($\alpha_{nm} = 0$). The new algorithm can cause divergence of a thermodynamic quantity at high densities. Another example of non detailed balance sampling would be when several trial moves (such as displacement, volume changes (required for an NPT ensemble) and molecule rotations) are performed in a predetermined sequence. Once again, the reserve move would be impossible, which leads to wrong sampling. Such a sequential scheme, however, can be applied when choosing particles. Instead of randomly choosing a particle in a system, an order in choosing the particles one after the other can be established for faster convergence of the simulation. Although this algorithm certainly breaks the microscopic reversibility condition, it has been shown to result in correct MC sampling that satisfies the limiting distribution equation (Equation 5.15).[75]

5.4.2 Configurational bias Monte Carlo technique

Standard MC methods sample the configurational phase space by performing rotational and translational changes on the molecules in the simulation box. They, however, do not account for the correct distribution of molecular conformations, in which molecules appear with the right statistical weights for different molecular structures. Achieving conformational equilibrium is an essential part of the internal equilibrium of a molecular fluid. Long chain molecules are particularly vulnerable to conformational changes, the number increasing with the molecule length and temperature. The practical solution of sampling the conformational phase space with the correct distribution was first suggested by Rosenbluth and Rosenbluth[36] and later extended by Siepmann and Frenkel.[37] The method is known as the Configurational Bias Monte Carlo (CBMC). Conformational changes are obtained by regrowing the molecule or part of it bead by bead

such that more probable configurations occur with a higher frequency. Each bead (or monomeric unit) is part of the molecule and can be considered as a single potential site interacting with its neighbors through intra and inter-molecular forces. An accurate model of long chain molecules usually comprises stretching, bending and torsional intra-molecular plus external non-bonded interactions (Figure 5.6).

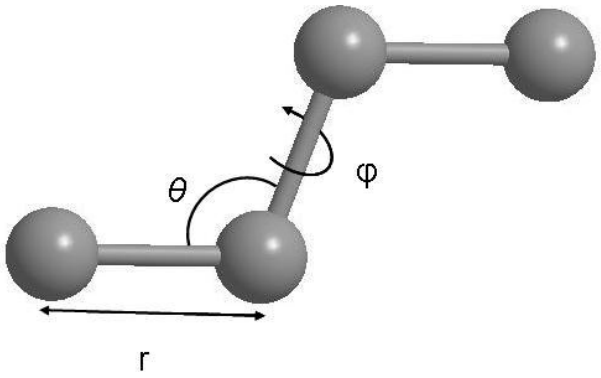


Figure 5.6. Segment of a long chain molecule with the intra-molecular degrees of freedom (stretching (r), bending (θ) and torsional changes (ϕ)). Each ball corresponds to a single potential site (bead).

The generation of a new conformation starts by selecting a molecule and a bead from it at random and removing the remaining of the molecule at either side of the bead. The removed segment will be kept in the computer memory in order to be compared with the newly regrown one. The regrowth starts with placing a new bead around the old position, first by considering the intramolecular interactions for the model. The bonded energy can be written as a sum of the vibrational, bending and torsional potentials ($U_{vib}(r),\ U_{bend}(\theta),\ U_{vib}(\phi)$):

$$U_{bond}(r,\theta,\phi) = U_{vib}(r) \ + \ U_{bend}(\theta) \ + \ U_{tors}(\phi) \tag{5.28}$$

The length of the bond is often subjected to a harmonic potential with oscillations around

an equilibrium distance. Thus the trial orientations can be distributed on an outer portion of a sphere, enclosed by two spherical surfaces with radii that are slightly bigger and slightly smaller than the equilibrium bond distance. Only a limited number of k trials are generated in order to speed up the simulation. The probability of generating a new position n from this set of trials configuration $\mathbf{r}$ is:

$$\rho_{n,bond}\mathrm{d}\mathbf{r} = \frac{\exp\left[-u_{n,bond}(\mathbf{r})/k_BT\right]\mathrm{d}\mathbf{r}}{\int \mathrm{d}\mathbf{r}\,\exp\left[-u_{i,bond}(\mathbf{r})/k_BT\right]} = \frac{\exp\left[-u_{n,bond}(\mathbf{r})/k_BT\right]\mathrm{d}\mathbf{r}}{\mathrm{C}} \tag{5.29}$$

where $u_{n,bond}$ is the bonded energy of monomer n, and C is the normalization constant related to the bonded energy. Since r, θ and ϕ are used to represent the conformations of the molecule (see Figure 5.6), it is convenient to work with spherical coordinates:

$$\mathrm{d}\mathbf{r} = r^2\mathrm{d}r\;\mathrm{d}cos\theta\;\mathrm{d}\phi \tag{5.30}$$

The probability that this new trial position is generated then becomes:

$$\rho_{n,bond}\mathrm{d}\mathbf{r} = \frac{\exp\left[-u_{n,vib}(r)/k_BT\right]\exp\left[-u_{n,bend}(\theta)/k_BT\right]\exp\left[-u_{n,tors}(\phi)/k_BT\right]r^2\mathrm{d}r\;\mathrm{d}cos\theta\;\mathrm{d}\phi}{\mathrm{C}} \tag{5.31}$$

A new trial is obtained, by generating a random vector within the portion of a sphere limited by the two radii and calculating the distance (r), bending (θ) and torsional (ϕ) angles. The distance and the angles are accepted with a probability according to equation 5.31. If rejected, another vector is generated until one gets accepted. In case the new position corresponds to the second bead in the molecule, any random vector is accepted

with the probability distribution according to the vibrational energy, since no bending or torsional restraints are present. The probability of generating the third bead is also simplified, given that no torsional changes are present.

$$\rho_{n,bond}\mathrm{d}\mathbf{r} = \frac{\exp\left[-u_{n,vib}(r)/k_BT\right]\exp\left[-u_{n,bend}(\theta)/k_BT\right]r^2\mathrm{d}r\ \mathrm{d}cos\theta\ \mathrm{d}\phi}{\mathrm{C}} \tag{5.32}$$

The next step is to calculate the external, non-bonded energy probability of the selected bead. At this point, selecting the new position depends on the interactions with the beads from the other molecules around and those from the same molecule separated by more than 3 monomeric units:

$$\rho_{n,ext}(\mathbf{r}_n) = \frac{\exp\left[-u_{n,ext}(\mathbf{r}_n)/k_BT\right]}{\sum_{i=1}^{k}\exp\left[-u_{i,ext}(\mathbf{r}_n)/k_BT\right]} \tag{5.33}$$

where $u_{n,ext}(\mathbf{r}_n)$ is the external energy for the new bead and the denominator represents its external configurational partition function. The latter term is given the symbol $w_{i,ext}(n)$:

$$\rho_{n,ext}(\mathbf{r}_n) = \frac{\exp\left[-u_{n,ext}(\mathbf{r}_n)/k_BT\right]}{w_{i,ext}(n)} \tag{5.34}$$

Once the entire chain is regrown, the Rosenbluth factor is calculated:

$$W_{ext}(n) = \prod_{i=1}^{l} w_{i,ext}(n) \tag{5.35}$$

where, l is the number of beads regrown.

In order to compare the thus obtained new conformation to the old (o) one, it is necessary to repeat the same steps for the old chain. This time, the first step of evaluating

the bonded probability is omitted as it does not influence the final acceptance criterion for accepting or rejecting the new structure. Each ”new” bead is also accepted with 100 % probability as their positions in the old chain are predetermined. Same number of different trials around the old monomeric unit however is generated, calculating the external energy for each one ($u_{i,ext}(o)$) and the external configurational partition function:

$$w_{i,ext}(o) = \sum_{i=1}^{k} \exp\left[-u_{i,ext}(\mathbf{r}_o)/k_B T\right] \tag{5.36}$$

When the old conformation is retraced, once again the Rosenbluth factor is calculated:

$$W_{ext}(o) = \prod_{i=1}^{l} w_{i,ext}(o) \tag{5.37}$$

The correct sampling of the transition from the old to the new chain conformation is guaranteed by the detailed balance condition (Equation 5.17). In what follows, this condition is discussed in details, comparing a single old and a new bead as a conformational change, the extension being straightforward for a larger segment of the molecule. The difference with respect to the conventional MC scheme is that the transition matrix probability is biased according to the above described method of generating a new bead from the old one. It depends on the product of generating a trial orientation according to the bonded probability (Equation 5.29) and the external energy probability of the selected bead (Equation 5.33):

$$\alpha_{on} = \rho_{i,bond}(n)\rho_{i,ext}(n)\alpha'_{on} \tag{5.38}$$

where α'_{on} is a symmetric transition matrix, such as the transition matrix in a conventional

MC scheme. The detailed balance condition is:

$$\rho_o \alpha_{on} acc_{on} = \rho_n \alpha_{no} acc_{no} \tag{5.39}$$

By replacing α_{on} with the expression in Equation 5.38, it becomes:

$$\rho_o \rho_{bond}(n) \rho_{ext}(n) \alpha'_{on} acc_{on} = \rho_n \rho_{bond}(o) \rho_{ext}(o) \alpha'_{no} acc_{no} \tag{5.40}$$

Since α' is symmetric and $\rho_o \propto \exp[-u(o)/k_BT]$, the left and right side of the equation can be written as follows:

1. Left side

$$\exp[-u(o)/k_BT] \frac{\exp[-u_{bond}(n)/k_BT]}{\mathrm{C}} \frac{\exp[-u_{ext}(n)/k_BT]}{w_{ext}(n)} acc_{on} \tag{5.41}$$

2. Right side

$$\exp[-u(n)/k_BT] \frac{\exp[-u_{bond}(o)/k_BT]}{\mathrm{C}} \frac{\exp[-u_{ext}(o)/k_BT]}{w_{ext}(o)}) acc_{no} \tag{5.42}$$

It is of note that since $u(o) = u_{bond}(o) + u_{ext}(o)$ one can express the probability of being in the old configuration as:

$$\exp[-u(o)/k_BT] = \exp[-u_{bond}(o)/k_BT] \ \exp[-u_{ext}(o)/k_BT] \tag{5.43}$$

Using the above expression, the detailed balance condition can be further simplified; having the same bonding configurational partition function for the old and new bead, the

final acceptance criterion becomes:

$$acc_{on} = \min\left\{1, \frac{w_{ext}(n)}{w_{ext}(o)}\right\} \tag{5.44}$$

The move is accepted based solely on the external non-bonded energy of the trials for the new and old bead. The convenience of using the Rosenbluth scheme in the off-lattice case comes from the fact that bonded energy is omitted in the final acceptance rule, which decreases the computational time. If more than one bead is grown, expressions 1 and 2 would grow with additional terms, corresponding to each bead in the regrown and its counterpart old conformation. The acceptance criterion can be simplified again to give this time a product of the w_{ext} factors for each new and old bead, which are the two Rosenbluth factors (Equations 5.35 and 5.37).

$$acc_{on} = \min\left\{1, \frac{W_{ext}(n)}{W_{ext}(o)}\right\} \tag{5.45}$$

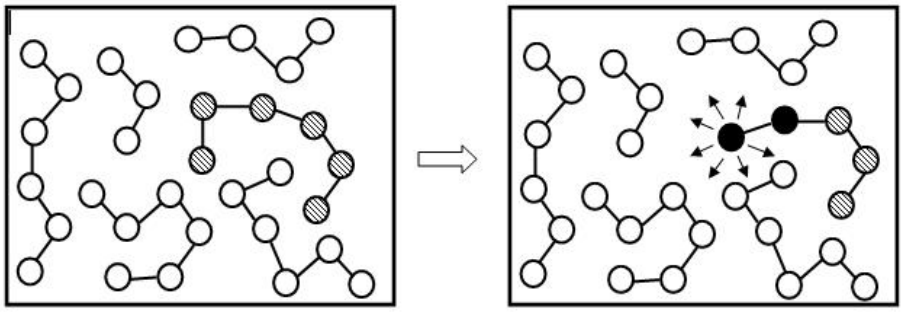


Figure 5.7. Schematic representation of the CBMC algorithm in 2D. The black beads correspond to part of the molecule which is regrown with the arrows indicating the possible orientations for the next bead.

The CBMC technique is tightly coupled with the Gibbs Ensemble Monte Carlo (GEMC) method, especially in the construction of phase diagrams of alkane chains. The

particle exchange moves are essential to attain equilibrium in the Gibbs Ensemble. The insertion of a chain molecule in the liquid phase must be accompanied with a change in the conformational structure of the molecule in order to avoid overlapping and therefore it is inserted bead by bead, in exactly the same manner as when a chain is regrown. The particle exchange move, however, still presents a challenge for long molecules with complex architecture. The configurational bias technique has to be implemented not only for the particle exchange, but also with the MC move inside each of the simulation boxes, for the regrowth (conformational change) of the chain.

Generation of trial positions is time consuming, when a branched chain is considered. The acceptance of a trial configuration of the branched site now depends on the bond length, the torsional angle and the two angles between the bond vector and the two existing branches. The acceptance probability is thus decreased and the efficiency for successful generation of a branched site is about 2-10 times lower than for a linear molecule. This method was later improved by Nath and Khare[67] by growing all the atoms from a branch point simultaneously instead of one after another. This technique, described in details by Macedonia and Maginn[66] was found to be faster with an efficiency of generating new configurations of branched sites an order of magnitude higher for simple molecules and increases as the molecule becomes more complex. Yet, the acceptance probability of the inserted molecule will depend on number of the preliminary established configurations of a single branched segment.

CHAPTER VI

CONCLUSION

The Wang-Landau method can be applied to effectively calculate vapor-liquid equilibria curves. It is simple to use since a single simulation run can give rise to equilibrium properties of the system at a given temperature. In this respect, it retains the simplicity of GEMC and transition matrix MC schemes. In addition, efficient sampling of the phase space is achieved through the incorporation of configurational bias Monte Carlo or hybrid Monte Carlo techniques. Their integration significantly reduces the computational time when sampling dense liquid regions. The integration of MD trajectories, within the HMC algorithm, helps for the faster convergence of the simulation, due to the concerted and deterministic moves of the particles. Furthermore, the implementation of the WL algorithm in the NPT ensemble does not rely on biasing strategies for the efficient insertion of particles in the dense liquid regions. Thus sampling of fluids with a variety of molecular architecture is possible, without the need to adjust the configurational bias according to the structural characteristics of the molecule. Sampling of the liquid regions with the HMC method also removes the necessity of adjusting the biasing strategy to fit the molecular structure, which is an issue present in the Gibbs ensemble or grand canonical ensemble.

The WL scheme thoroughly samples the ensemble microstates of vapor and liquid phases at coexistences. A noticeable feature of the method is the sampling of the interface

region, which is avoided in other methods. This is achieved through an accurate estimation of the ensemble partition function. The latter, in turn, is obtained by evaluating a biasing function through an iterative scheme. In the NPT ensemble the convergence of the simulation can be easily verified by ensuring the smoothness and continuity of the canonical partition function. The simulation is conducted without prior precise knowledge of equilibrium properties of the system - even though the simulations are performed at a constant pressure, prior knowledge of the saturation pressure is not needed.

Wang-Landau simulations in the NPT ensemble is an efficient and reliable method for simulating vapor and liquid phases in coexistence. Critical properties of copper from our simulations have shown to be within the experimental range available and extrapolation line of the experimental saturation pressures lies close to the simulated ones. Vapor and liquid densities of linear and branched alkanes are in excellent agreement with experimental data and critical points of isobutane and isopentane being within 1 to 7 % of the available experimental data, while those of *n*-alkanes are within 2 to 3 %. The accuracy of the simulation results proves the reliability of the method.

REFERENCES

1. Panagiotopoulos, A. Z. Phase coexistence properties of liquid mercury: A simulation study. *Mol. Phys.* **1987**, *61*, 814.

2. Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, 2nd ed.; Academic Press: San Diego, CA, 2001.

3. Wang, F.; Landau, D. P. Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. *Phys. Rev. E* **2001**, *64*, 056101/1–056101/16.

4. Wang, F.; Landau, D. P. Efficient, Multiple-Range RandomWalk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.* **2001**, *86*, 2050–2053.

5. Landau, D. P.; Tsai, S. .; Exler, M. A new approach to Monte Carlo simulations in statistical physics: Wang-Landau sampling. *Am. J. Phys.* **2004**, *72*, 1294–1302.

6. Schulz, B. J.; Binder, K.; Mller, M.; Landau, D. P. Avoiding boundary effects in Wang-Landau sampling. *Phys. Rev. E* **2003**, *67*, 067102/1–067102/2.

7. Schulz, B. J.; Binder, K.; Mller, M. Flat histogram method of Wang-Landau and N-fold way. *International Journal of Modern Physics C* **2002**, *13*, 477–494.

8. Rathore, N.; De Pablo, J. J. Monte Carlo simulation of proteins through a random walk in energy space. *J. Chem. Phys.* **2002**, *116*, 7225–7230.

9. Rubio, A. M.; Storey, M.; Lodge, J. F. M.; Freire, J. J. Dynamics of bond-fluctuation model chains in good and theta solvents. *Macromol. Theory Simul.* **2002**, *11*, 171–183.

10. Rampf, F.; Binder, K.; Paul, W. The phase diagram of a single polymer chain: New insights from a new simulation method. *Journal of Polymer Science, Part B: Polymer Physics* **2006**, *44*, 2542–2555.

11. Shell, M. S.; Debenedetti, P. G.; Panagiotopoulos, A. Z. Generalization of the Wang-Landau method for off-lattice simulations. *Phys. Rev. E* **2002**, *66*, 056703/1–056703/9.

12. Mastny, E. A.; De Pablo, J. J. Direct calculation of solid-liquid equilibria from density-of-states Monte Carlo simulations. *J. Chem. Phys.* **2005**, *122*, 1–6.

13. Virnau, P.; Mller, M.; MacDowell, L. G.; Binder, K. Phase behavior of n-alkanes in supercritical solution: A Monte Carlo study. *J. Chem. Phys.* **2004**, *121*, 2169–2179.

14. Rathore, N.; Knotts IV, T. A.; De Pablo, J. J. Density of states simulations of proteins. *J. Chem. Phys.* **2003**, *118*, 4285–4290.

15. Errington, J. R. Direct calculation of liquid-vapor phase equilibria from transition matrix Monte Carlo simulation. *J. Chem. Phys.* **2003**, *118*, 9915–9925.

16. Desgranges, C.; Delhommelle, J. Phase equilibria of molecular fluids via hybrid Monte Carlo Wang-Landau simulations: Applications to benzene and n -alkanes. *J. Chem. Phys.* **2009**, *130*, 244109.

17. Aleksandrov, T.; Desgranges, C.; Delhommelle, J. Vapor-liquid equilibria of copper using hybrid Monte Carlo Wang–Landau simulations. *Fluid Phase Equilibria.* **2010**, *287*, 79–83.

18. Stull, D.; Gray, D. *American Institute of Physics Handbook*, 3rd ed.; McGraw Hill: New York, 1972.

19. Linde, D. *CRC Handbook of Chenistry and Physics*, 2nd ed.; CRC Press Inc: Cleveland, OH, 2003.

20. Gelb, L. D.; Chakraborty, S. N. Boiling point determination using adiabatic Gibbs ensemble Monte Carlo simulations: Application to metals described by embedded-atom potentials. *J. Chem. Phys.* **2011**, *135*, 224113–1–7.

21. Brass, A.; Pendleton, B. J.; Chen, Y.; Robson, B. Hybrid Monte Carlo simulations theory and initial comparison with molecular dynamics. *Biopolymers* **1993**, *33*, 1307–1315.

22. Vargaftik, N.; Vinoradov, Y.; Yargin, V. *Handbook of Physical Properties of Liquids and Gases*; Begell House: New York, 1996.

23. Nath, S. K.; De Pablo, J. J. Simulation of vapour-liquid equilibria for branched alkanes. *Molecular Physics* **2000**, *98*, 231–238.

24. Martynyuk, M. M.; Panteleichuk, O. G. Critical temperature of metals by the method of the electric explosion of wires under pressure. *High Temperature* **1976**, *14*, 1075–1079.

25. Luo, S. N.; Ahrens, T. J.; Cagin, T. A.; Strachan, W. A.; Goddard, W. A.; Swift, D. C. Maximum superheating and undercooling Systematics, molecular dynamics simulations, and dynamic experiments. *Phys. Rev. B* **2003**, *68*, 134206–1–11.

26. Pottlacher, G.; Jager, H. A review of determinations of critical point data of metals using subsecond pulse heating techniques. *J. Non-Crys. Sol.* **1996**, *205-207*, 265–269.

27. Gathers, G. R. Dynamic methods for investigating thermophysical properties of matter at very high temperatures and pressures. *Rep. Prog. Phys.* **1986**, *49*, 341–396.

28. Young, A. B., D.A. Critical point of metals from the van der Waals model. *Phys. Rev. A* **1971**, *3*, 364–371.

29. Blairs, S.; Abbasi, M. H. Correlation between surface tension and critical temperatures of liquid metals. *J. Colloid Interface Sci.* **2006**, *304*, 549–553.

30. Likalter, A. A. Phase coexistence properties of liquid mercury: A simulation study. *Phys. ReV. B* **1996**, *53*, 4386–4392.

31. Nikitin, E. D.; Pavlov, P. A.; Popov, A. P. Vapour-liquid critical temperatures and pressures of normal alkanes with from 19 to 36 carbon atoms, naphthalene and m-terphenyl determined by the pulse-heating technique. *Fluid Phase Equilibria* **1997**, *141*, 155–164.

32. VonNiederhausern, D. M.; Wilson, G. M.; Giles, N. F. Critical point and vapor pressure measurements at high temperatures by means of a new apparatus with ultralow residence times. *J. Chem. Engin. Data* **2000**, *45*, 157–160.

33. Martin, M. G.; Siepmann, J. I. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569–2577.

34. Errington, J. R.; Panagiotopoulos, A. Z. A new intermolecular potential model for the n-alkane homologous series. *J. Phys. Chem. B* **1999**, *103*, 6314–6322.

35. Ungerer, P.; Beauvais, C.; Delhommelle, J.; Boutin, A.; Rousseau, B.; Fuchs, A. H. Optimization of the anisotropic united atoms intermolecular potential for n-alkanes. *J. Chem. Phys.* **2000**, *112*, 5499–5510.

36. Rosenbluth, M. N.; Rosenbluth, A. W. Monte carlo calculation of the average extension of molecular chains. *J. Chem. Phys.* **1955**, *23*, 356–359.

37. Siepmann, J. I.; Frenkel, D. Configurational bias Monte Carlo: A new sampling scheme for flexible chains. *Molec.Phys.* **1992**, *75*, 59–70.

38. Cui, S. T.; Cummings, P. T.; Cochran, H. D. Configurational bias Gibbs ensemble Monte Carlo simulation of vapor-liquid equilibria of linear and short-branched alkanes. *Fluid Phase Equilibria* **1997**, *141*, 45–61.

39. Zhuravlev, N. D.; Siepmann, J. I. Exploration of the vapour-liquid phase equilibria and critical points of triacontane isomers. *Fluid Phase Equilibria* **1997**, *134*, 55–61.

40. Zhuravlev, N. D.; Siepmann, J. I.; Schure, M. R. Surface coverages of bonded-phase ligands on silica: A computational study. *Anal. Chem.* **2001**, *73*, 4006–4011.

41. Neubauer, B.; Delhommelle, J.; Boutin, A.; Tavitian, B.; Fuchs, A. H. Monte Carlo simulations of squalane in the Gibbs ensemble. *Fluid Phase Equilibria* **1999**, *155*, 167–176.

42. Fitzgerald, M.; Picard, R. R.; Silver, R. N. Canonical transition probabilities for adaptive Metropolis simulation. *Europhys. Lett.* **1999**, *46*, 282–287.

43. Fitzgerald, M.; Picard, R. R.; Silver, R. N. Monte Carlo transition dynamics and variance reduction. *J. Stat. Phys.* **2000**, *98*, 321–345.

44. Ferrenberg, A. M.; Swendsen, R. H. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* **1988**, *61*, 2635–2638.

45. Ganzenmüller, G.; Camp, P. J. Applications of Wang-Landau sampling to determine phase equilibria in complex fluids. *J. Chem. Phys.* **2007**, *127*, 154504–1–10.

46. Desgranges, C.; Kastl, E. A.; Aleksandrov, T.; Delhommelle, J. Optimisation of multiple time-step hybrid Monte Carlo Wang-Landau simulations in the isobaric-isothermal ensemble for the determination of phase equilibria. *Molec. Sim.* **2010**, *36*.

47. Berg, B. A.; Neuhaus, T. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett.* **1992**, *68*, 9–12.

48. Beale, P. D. Exact distribution of energies in the two-dimensional ising model. *Phys. Rev. Lett.* **1996**, *76*, 78–81.

49. Bortz, A. B.; Kalos, M. H.; Lebowitz, J. L. A new algorithm for Monte Carlo simulation of Ising spin systems. *J. Comput. Phys.* **1975**, *17*, 10–18.

50. Binder, K.; Landau, D. P.; Ferrenberg, A. M. Thin Ising films with competing walls: A Monte Carlo study. *Phys. Rev. E* **1995**, *51*, 2823–2838.

51. Jain, T. S.; De Pablo, J. J. A biased Monte Carlo technique for calculation of the density of states of polymer films. *J. Chem. Phys.* **2002**, *116*, 7238–7243.

52. Kumar, S. K.; Szleifer, I.; Panagiotopoulos, A. Z. Determination of the chemical potentials of polymeric systems from Monte Carlo simulations. *Phys. Rev. Lett.* **1991**, *66*, 2935–2938.

53. Yan, Q.; Faller, R.; De Pablo, J. J. Density-of-states Monte Carlo method for simulation of fluids. *J. Chem. Phys.* **2002**, *116*, 8745–8749.

54. Rathore, N.; Knotts IV, T. A.; De Pablo, J. J. Erratum: Density of states simulations of proteins (J. Chem. Phys. (2003) 118 (4285)). *J. Chem. Phys.* **2003**, *118*, 9460–9461.

55. Bhatt, D.; Jasper, A. W.; Schultz, N. E.; Siepmann, J. I.; Truhlar, D. G. Critical Properties of Aluminum. *J. Am. Chem. Soc.* **2006**, *128*, 4224–4225.

56. Singh, J. K.; Adhikari, J.; Kwak, S. K. Vapor-liquid phase coexistence curves for Morse fluids. *Fluid Phase Equilibria* **2006**, *248*, 1–6.

57. Cheng, C.; Xu, X. Molecular Dynamics Calculation of Critical Point of Nickel. *Int. J. Thermophys.* **2005**, *28*, 9–19.

58. Bretonnet, J.-L.; Bomont., J.-M. Phase coexistence properties of liquid mercury: A simulation study. *Rare Metals* **2006**, *25*, 567–571.

59. Xu, P.; Cagin, T.; Goddard, W. Assessment of phenomenological models for viscosity of liquids based on nonequilibrium atomistic simulations of copper. *J. Chem. Phys.* **2005**, *123*, 104506–1–8.

60. Desgranges, C.; Delhommelle, J. Shear viscosity of liquid copper at experimentally accessible shear rates:Application of the transient-time correlation function formalism. *J. Chem. Phys.* **2008**, *128*, 084506.

61. Panagiotopoulos, A. Z. Molecular simulation of phase coexistence: Finite-size effects and determination of critical parameters for two- and three-dimensional Lennard-Jones fluids. *Int. J. Thermophys.* **1994**, *15*, 1057–1072.

62. Hess, H. Critical data and vapor pressures for aluminium and copper. *Z. Metallkd.* **1998**, *89*, 388–393.

63. Duane, S.; Kennedy, A. D.; Pendleton, B. J.; Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **1987**, *195*, 216–222.

64. Forrest, B. M.; Suter, U. W. Hybrid Monte Carlo simulations of dense polymer systems. *The Journal of chemical physics* **1994**, *101*, 2616–2629.

65. Mehlig, B.; Heermann, D. W.; Forrest, B. M. Hybrid Monte Carlo method for condensed-matter systems. *Phys. Rev. B* **1992**, *45*, 679–685.

66. Macedonia, M. D.; Maginn, E. J. A biased grand canonical Monte Carlo method for simulating adsorption using all-atom and branched united atom models. *Mol. Phys.* **1999**, *96*, 1375–1390.

67. Nath, S. K.; Khare, R. New forcefield parameters for branched hydrocarbons. *J. Chem. Phys.* **2001**, *115*, 10837–10844.

68. Wescott, J. T.; Kung, P.; Nath, S. K. Vapour-liquid coexistence properties and critical points of two branched alkanes series. *Fluid Phase Equilibria* **2003**, *208*, 123–139.

69. Nath, S. K.; Escobedo, F. A.; De Pablo, J. J. On the simulation of vapor-liquid equilibria for alkanes. *J. Chem. Phys.* **1998**, *108*, 9905–9911.

70. Allen, M.; Tildesley, D. *Computer Simulation of Liquids*, 1st ed.; Oxford University Press: Oxford, 1987.

71. Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **1996**, *87*, 1117–1157.

72. Trotter, H. On the product of semi-groups of operators. *Proc. Amer. Math. Soc.* **1959**, *10*, 545–551.

73. Weast, R.; Grasselli, J. *Handbook of Data on Organic Compounds*, 2nd ed.; CRC Press Inc: Boca Raton, FL, 1989.

74. Kudchaker, A. P.; Zwolinski, B. J. Vapor pressures and boiling of normal alkanes, C_{21} to C_{100}. *J. Chem. Eng. Data* **1966**, *11*, 253.

75. Manousiouthakis, V. I.; Deem, M. W. Strict detailed balance is unnecessary in Monte Carlo simulation. *J. Chem. Phys.* **1999**, *110*, 2753–2756.