Determination of Phase Equilibria for Long-Chain Linear Hydrocarbons by Monte Carlo Simulation

by

Nicholas Bruce du Preez

[BSc.(Eng.)]

University of Natal, Durban

Submitted in fulfillment of the academic requirements for the degree of Master of Science in Engineering at the School of Chemical Engineering, University of KwaZulu-Natal

Durban

2005

The Fly

Little Fly,
Thy summer's play
My thoughtless hand
Has brushed away.

Am not I
A fly like thee?
Or art not thou
A man like me?

For I dance
And drink and sing,
Till some blind hand
Shall brush my wing.

If thought is life
And strength and breath,
And the want
Of thought is death,

Then am I
A happy fly,
If I live
Or if I die.

Songs of Experience
William Blake (1757–1827)

Preface

The work presented in this dissertation was performed at the University of KwaZulu-Natal, Howard Campus, from February 2003 to December 2004. This study was supervised by Deresh Ramjugernath (University of KwaZulu-Natal) and Kim Bolton (University of Borås).

This dissertation is submitted as the full requirement for the degree Master of Science in Chemical Engineering. All of the work presented in this dissertation is original, unless otherwise stated. It has not (in whole or in part) been previously submitted to any tertiary institute as part of a degree.

Nicholas du Preez

21 December 2005

Acknowledgments

I would like to thank the following for all the input they have given during this study:

- My supervisors, Deresh Ramjugernath (University of KwaZulu-Natal) and Kim Bolton (University of Borås) for their support and guidance during this study.
- Tyrone McKnight (ChemEng, UKZN) whose help and advice during this study has been very much needed and appreciated!!
- Marcus Martin (Sandia National Laboratories) for all his help with the TOWHEE code and CBMC methods.
- Stefan Scriba (ElecEng, UKZN) for all his help with LaTeX 2_ε.
- Yash Nannoolal (ChemEng, UKZN) for his help with data collection from the DDB.
- Anders Askåsen and Peter Ahlström from University of Borås for their help during this study.
- National Supercomputer Centre (Linköping University, Sweden) for computer time on Monolith and SGI3K.
- Graduate Assistantship from the University of KwaZulu-Natal and National Research Fund Thuthuka Programme for their financial assistance over the last two years.
- My friends and colleagues of ChemEng, in particular Kerri, John-Roy, Scott and Roger.
- My family for their continued support throughout my long years of study.

Abstract

The focus of this study was to determine the coexistence phase equilibria for three groups of long-chain linear hydrocarbons (n-alkanes, 1-alkenes and 1-alcohols) using Monte Carlo simulation. Three common transferable united-atom force fields were used in the simulations: OPLS-UA (Jorgensen et al., 1984), TraPPE-UA (Martin and Siepmann, 1998) and NERD (Nath, Escobedo, de Pablo and Patramai, 1998). Isothermal phase equilibria was calculated over a temperature range from approximately the normal boiling point up to just below the critical temperature. The liquid and vapour densities and vapour pressures were determined from the simulations. The density results were then fitted using least-squares regression to the scaling law and the law of rectilinear diameters in order to estimate the critical properties. The vapour pressure data were fitted using least-squares to the Clausius-Clapeyron equation to estimate the normal boiling points.

The NVT-Gibbs ensemble method was used to simulate the pure-component coexistence of the vapour and liquid phases. The NPT-Gibbs ensemble was used to simulate the n-alkane binary mixtures. Two forms of configurational-bias Monte Carlo (standard CBMC and coupled-decoupled CBMC) were used to increase the number of swap moves accepted during the simulations. Dual-cutoff CBMC was implemented with a second cut-off of 5Å in order to speed up the CBMC calculations. Minimum image and a spherical potential truncation after 14Å were implemented with standard tail corrections. BIGMAC and TOWHEE were the two Fortran-77 codes used to simulate the hydrocarbon compounds. BIGMAC was used in the simulations of non-polar molecules and TOWHEE was used in the simulations of polar molecules. System sizes ranged from 300 (for the C_8's) down to 100 molecules (for the C_{20}'s). The simulations were typically equilibrated for at least 30000 cycles and production runs ranged from 50000 to 120000 cycles for the different hydrocarbon groups. Standard deviations of the calculated thermophysical properties were between 1–3% for the liquid densities and 10–20% for the

vapour densities and vapour pressures.

It was found that the coexistence density curves were generally in good agreement with experiment for all the hydrocarbon groups investigated (the OPLS-UA force field being the exception). The chain-length appeared to have little effect on the quality of the calculated thermophysical properties. The chain-length did however increase the time required to perform the simulations substantially. The vapour pressures were consistently over-predicted by NERD and TraPPE-UA. The normal boiling points were typically under-predicted by 2–5%. The critical temperatures and densities were predicted to within 1–5% of experimental values. The n-alkane mixtures were satisfactorily predicted using the NPT-Gibbs ensemble. While both the NERD and TraPPE-UA force fields were shown to be substantially more accurate compared to the OPLS-UA force field, there was little difference between their predictions. Thus, it is likely that the added complexity of using the bond-stretching potential (used by NERD) is unnecessary. The results of this study show that Monte Carlo simulation may be used to predict vapour-liquid coexistence properties of long-chain hydrocarbons and to approximate critical properties. However, current force fields require more refinement in order to accurately predict the hydrocarbon thermophysical properties. Plus, faster computing speeds are required before Monte Carlo simulation becomes an industrially viable method.

Table of Contents

List of Figures

List of Tables

Nomenclature

Latin Alphabet Symbols

A	Helmholtz free energy
$\mathcal{A}$	dynamic variable
$\mathcal{B}$	dynamic variable
C	y-int of Clausius-Clapeyron plot regression
C'	gradient of Clausius-Clapeyron plot regression
E	internal energy
$\mathcal{H}(\mathbf{p}, \mathbf{r})$	Hamiltonian
$K(o \rightarrow n)$	flow of configurations from o to n
$\mathcal{K}$	kinetic energy
L	box length
N	number of particles or molecules
N_{sys}	number of systems
P	pressure
Q	canonical partition function
Q_G	Gibbs ensemble partition function
R	Universal gas constant
S	entropy
T	temperature
$\mathcal{U}$	potential energy
V	volume
W	Rosenbluth weight

$\alpha(o \rightarrow n)$	probability of generating state n from o
β	reciprocal temperature, $1/(k_B T)$
γ	critical component
ϵ	characteristic energy in pair potential
ϵ_0	permittivity of free space
ϵ_s	relative permissibility
θ	bond bending angle
μ	chemical potential
$\pi(o \rightarrow n)$	transition probability from o to n
ρ	density
σ	characteristic distance in a pair potential
ϕ	dihedral angle
ω	dynamic variable

Subscript and Superscript Symbols

b	boiling point
c	critical property
ex	excess property
i	i^{th} component
j	j^{th} component
test	test particle (or molecule)
therm	thermodynamic
vir	virial

Other Symbols

$< \cdots >$	property average
$acc(o \rightarrow n)$	acceptance probability of a move from o to n
ΔH_{vap}	heat of vapourisation
$\wp$	probability density
NBP	normal boiling point

Chapter 1

Introduction

This study is concerned with the determination of phase equilibria by Monte Carlo simulation for long-chain linear hydrocarbons using three common transferable united-atom force fields. These force fields, OPLS-UA (Jorgensen, 1983), TraPPE-UA (Martin and Siepmann, 1998) and NERD (Nath, Escobedo, de Pablo and Patramai, 1998) have been used to predict isothermal phase equilibria over temperature range from approximately the normal boiling point up to just below the critical temperature. The liquid and vapour densities and vapour pressures were determined from the simulation. These results have been interpolated to determine the normal boiling points and extrapolated to determine the critical properties.

The hydrocarbons investigated were selected from the sub-classes n-alkanes, 1-alkenes and 1-alcohols, with carbon numbers ranging from 8 up to ~30. For all discussions that follow, the term *hydrocarbons* will refer specifically to the above mentioned alkanes, alkenes and alcohols, unless otherwise stated.

These hydrocarbons were chosen for three reasons: first, the greatest amount of consistent experimental data are available for these hydrocarbons. The second reason is their simple molecular structure. Simulations involving long hydrocarbons are very time consuming and, by choosing a simpler molecular structure, one is able to make assumptions that speed up the simulations. These methods will be discussed later in Chapters 4 and 8. The third reason is that the physical chemistry of these hydrocarbons is better understood compared to many other long hydrocarbon groups and some experimental data is available. This allows for some comparison between theoretical predictions (calculated using Monte Carlo simulation) and available experimental observations.

Methods to predict the phase equilibria of long-chain hydrocarbons are of great importance. It is of particular use in the petrochemical industry, for the design of separation equipment. Unfortunately, much of the experimental data available for high molecular-weight hydrocarbons are not only scarce but often contradictory (Siepmann et al., 1993). There are many difficulties in the experimental determination of phase equilibria for long hydrocarbons. The greatest difficulty is that thermal instability of hydrocarbons occur from ~600K (Siepmann et al., 1993; Chen et al., 2001). This makes the measurement of critical properties and even normal boiling points, for the longer hydrocarbons, all but impossible with current experimental methods. The other reason, of lesser difficulty, is that the longer hydrocarbons are solid at room temperature. To date, a large effort has been put into methods such as equation-of-state, to predict these phase equilibria data. These methods often require critical properties, although there are recent efforts to use the normal boiling point as the reference temperature (Coniglio et al., 2000; Crampon et al., 2004; *DDB - Dortmund Data Bank*, 2004). Clearly the importance of determining accurate critical properties and normal boiling points cannot be understated. These experimental difficulties are avoided when using Monte Carlo simulation.

With molecular simulation, in general, all aspects of the simulation may be controlled. This allows one to investigate many difficult and important chemical systems. The use of simulation for the determination of critical properties of heavy hydrocarbons is one, but there are many other aspects of both practical and theoretical interest that molecular simulation may be used to investigate. For example, association in chemical systems (Chen and Siepmann, 2000; Chen and Siepmann, 2001) and solubilities of alkanes (Vlugt et al., 1999; Vlugt, 2000) or hydrogen (van den Berg et al., 2004) in zeolites to mention two other examples.

1.1 A Brief Overview

This section is a basic overview* of a number of the concepts used and discussed in this dissertation. It is only meant to be introductory in nature. Many of the topics presented here will be discussed in greater detail in later chapters.

*This section is partly based on the introductory review by Sadus (1999)

1.1.1 An Overview of Molecular Simulation

The two main methods in classical molecular simulations are Monte Carlo (MC) and molecular dynamics (MD) simulation methods. A good description of molecular simulation would be computational statistical mechanics. Using it one is able to determine macroscopic properties by evaluating theoretical models of molecular behaviour. This allows for the testing of theoretical models and the resolving of conflicts in experimental data (Siepmann et al., 1993). The first simulations were undertaken by Metropolis et al. (1953) where liquid simulations were performed on the MANIAC computer at Los Alamos. This was also the introduction of the MC simulation method. For MC simulations the intermolecular interactions are used to accept or reject trial configurations (microscopic states) which are generated randomly. Later, Alder and Wainwright (1957) introduced the MD simulation method. In MD simulations the equations of motion are used to predict the coordinate and momentum changes due to the intermolecular forces experienced by the molecules. The choice of method used depends on the properties that are of interest. Equilibrium data are the focus of this work and not time-dependent dynamic properties. For this reason the MC method is used since MC is time-independent. Of course MD simulations can be used for the determination of equilibrium data, but in the case of this study, MD is a much less desirable method. This is because large, polyatomic molecules are very calculation intensive and thus time consuming. In general, MD should not be used for determining equilibrium properties alone, rather only when time-dependent properties are also of interest (such as in drug design) should it be used.

1.1.2 The Monte Carlo Method

The MC method is a stochastic strategy that relies on probabilities. Trial system configurations are randomly generated and then selected based on an acceptance-rejection criteria, which is based on the calculation of the change in energy of the system from the current configuration too the trial one. The acceptance-rejection criteria is then compared with a random number, and if greater, then the trial configuration is selected. Of all the possible trial configurations, only a few make significant contributions to the configurational properties of the system. Thus, in order for MC simulations to be practical, some method is needed to generate these significant configurations. This is done by generating a Markov chain. A Markov chain is a sequence of trials that depend only on its

immediate predecessor. In this way, new configurations are accepted if they are more favorable than the previous configuration. Generally this means that the new configuration is of a lower energy than the previous but because the acceptance-rejection criterion is compared with a random number it is possible that an increase in energy will occur. As will be shown in a later chapter, all decreases in energy are accepted and the greater the increase in the energy the less likely it is to be accepted. The common methods of generating new configurations are displacing molecules within a system, swapping molecules between simulation boxes, the addition and removal of molecules from the system. A large research effort has been spent in attempts to optimize these methods and many others methods have since been developed, such as aggregative-volume-bias (Chen and Siepmann, 2000; Chen and Siepmann, 2001) and concerted rotation (Dodd et al., 1993).

1.1.3 Ensembles used with Monte Carlo

Metropolis et al. (1953) originally performed their simulations in the canonical ensemble, where the number of particles (N), temperature (T), and volume (V) are held constant. There are many possible ensembles, each describing different macroscopic conditions. The MC method has since been extended to the other common ensembles: isothermal-isobaric (NPT) by McDonald (1972); grand-canonical (μVT) by Valleau and Cohen (1980); and microcanonical (NVE) by Ray (1991).

1.1.4 Monatomic Molecules to Chain Molecules

Initially all simulations involved monatomic molecules or very simple poly-atomic molecules and their intermolecular interactions were described by either hard sphere or Lennard-Jones potentials. Even in considering these simple systems, there are a great many practical difficulties in simulating them. In a real experiment many moles ($N \sim 10^{23}$) of the molecules are used. Any attempt to simulate such a large system is clearly impossible. Only systems of a much smaller size ($N \sim 10^{2}$) may be considered. Thus some method is required to compensate or avoid the finite-size errors due to such small systems being considered. Periodic boundary conditions is the standard method used, in conjunction with the minimum image convention and the truncation of intermolecular potentials.

Turning ones consideration to chain molecules, which is the focus of this study, there are more difficulties involved in their simulation. Much work has been done on long n-alkanes (Siepmann et al., 1993; Smit et al., 1995; Martin and Siepmann, 1998; Nath,

Escobedo and de Pablo, 1998). There are two major difficulties, the first is the insertion of the these long-chained molecules. The chance of their insertion into dense phases decreases rapidly as the size of the chains increase. Configurational-bias Monte Carlo (CBMC) (Frenkel and Smit, 2002) based on the work of Rosenbluth and Rosenbluth (1955) has been introduced as a method to facilitate the simulation of these long-chains. More recent modifications to CBMC have been introduced to improve the method. These include the introduction of coupled-decoupled CBMC by Martin and Siepmann (1999) for the correct growth of branched molecules and Dual-cutoff CBMC (DC-CBMC) by Vlugt (2000) to speed up the calculation time required for CBMC. These methods have been used extensively in this study to make the simulation of the hydrocarbons in question more feasible. The second difficulty is the number of interaction sites for large molecules. If one considers all the hydrogens and carbons, there are potentially very many sites. The solution that has been used extensively in the molecular simulation field and in this study is known as united-atom potentials. Here the hydrogens are collapsed into the carbon atoms and one can see that this reduces the number of interaction sites substantially. This has been shown to be an adequate assumption for not very dense systems (Chen and Siepmann, 1999). However, there is some question as to whether the center-of-mass (or interaction site) of the united-atom should be at the center of the united-atom such as in the OPLS-UA (Jorgensen et al., 1984), TraPPE-UA (Martin and Siepmann, 1998) and NERD (Nath, Escobedo, de Pablo and Patramai, 1998) force fields or somewhere in between the united-atoms as in anisotropic potential models such as Toxvaerd (Toxvaerd, 1990). Anisotropic models have not been considered in this study since the added complexity has not been shown to yield better results.

1.1.5 Polar and Non-polar Molecules

The simulation of non-polar molecules are generally simplified with the assumption that molecules are completely neutral and only experience van der Waals forces. A long-ranged force is defined as one where the potential decays slower than r^{-d} (where d is the number of dimensions of the system). The van der Waals potential decays rapidly at r^{-6}, and in 3-dimensions, this means that it is a short-ranged force. Coulomb and dipole potentials decay with a much slower rate, r^{-1} and r^{-3} respectively. Thus, these are long-range forces and other methods must be employed to deal with these interactions. Many methods have been proposed, such as a reaction field or a particle-particle and

particle-mesh (PPPM) algorithm but the Ewald summation (Nosé and Klein, 1983) was the method chosen for this study.

1.1.6 Methods for Phase Equilibria

Two phases are in equilibrium if there is thermal, mechanical and chemical equilibrium. The first two are straight-forward enough to achieve with either the canonical or isothermal-isobaric ensembles but the third, chemical, is much more difficult to achieve. Also, if two phases are simulated in a single box there would be a phase interface which causes uncertainties. The Gibbs ensemble (Panagiotopoulos, 1987; Panagiotopoulos et al., 1988) method was introduced to solve these difficulties. With it, MC simulation as method of determining phase equilibria, came into its own. In this method the need to calculate the chemical potential in each box is avoided by using two separate (therefore no phase interface), but linked, simulation boxes. Although the Gibbs ensemble method has proved very successful and is now the standard method for determining phase equilibria, other methods have been developed. One method published shortly after the Gibbs method was Thermodynamic Integration (Kofke, 1993*a*; Kofke, 1993*b*). It requires an equilibrium point as a starting point and then the Gibbs-Duhem equation is integrated to trace the phase envelop. Another method, which is growing in popularity, is called Histogram Reweighting (Ferrenberg and Swendsen, 1988; Ferrenberg and Swendsen, 1989; Panagioropoulos et al., 1998). Energy vs Number-of-molecules histograms are collected from grand-canonical simulations and reweighted to different temperatures and chemical potentials to predict phase equilibria.

Chapter 2

Statistical Mechanics of Ensembles*

The main objective of statistical mechanics is to develop relationships which relate macroscopic properties, such as pressure or density, to the behaviour of the microscopic particles in a particular system. In this chapter, the fundamental postulates of ensemble statistical mechanics are presented, their application to the common ensembles and basic derivations of the thermodynamic averages are shown. These form the theoretical foundations for Monte Carlo simulations performed in this study.

2.1 Postulates of Statistical Mechanics

The general procedure when estimating macroscopic thermodynamic properties is to define postulates which can be used to evaluate these properties. The method commonly used is known as the 'ensemble method'. It was originally proposed by Gibbs, and is based on two postulates. These postulates cannot be justified *a priori*. They may only be 'proved' by comparing the results with experimentally determined values, and so far there is no evidence to doubt their validity.

2.1.1 Definition: The Ensemble

An ensemble is a collection (mental) of a large number of microstates ($N_{sys} \rightarrow \infty$) such that each microstate of the assembly is characterized by one or more extensive system variables (Kofke, 2003).

What this effectively amounts to is that all the systems or microstates within a particular ensemble have the same defining set of extensive and intensive variables (see

*This chapter is based largely on Allen and Tildesley (1987) and Rao (1994)

Table 2.1). There are a number of different possible ensembles which can be defined depending on the thermodynamic properties of interest. The most relevant ensembles will be briefly discussed later in this chapter.

2.1.2 First Postulate: Ergodicity

> The time average value of a mechanical variable (thermodynamic property) in the thermodynamic system is equal to the ensemble average value of the same variable subject to the condition that each of the systems in the ensemble is identical at a macroscopic level to the thermodynamic system of interest and the number of systems N_{sys} in the ensemble is extremely large, in the limit $N_{sys} \rightarrow \infty$.
>
> (Rao, 1994)

Or mathematically, for a given thermodynamic property $\mathcal{A}$:

$$\langle \mathcal{A} \rangle_{ens} = \langle \mathcal{A} \rangle_{time} = \lim_{t \to \infty} \frac{1}{t} \int_0^t \mathcal{A} \, d\tau \tag{2.1}$$

This postulate removes the requirement of determining the time average values for the thermodynamic properties. Instead it allows the use of ensemble average values for the determination of thermodynamic properties. This simplifies the determination of mechanical variables substantially. Unfortunately, there is still the practical difficulty of determining this ensemble average value. Without information on the probability of a particular state occurring and total possible states we cannot determine the ensemble average. This is where the second postulate is of use.

2.1.3 Second Postulate: The principle of equal *a priori* probabilities

> In an ensemble of isolated thermodynamic systems (microcanonical ensemble), the systems of the ensemble are distributed with equal probability over all quantum states which are consistent with the macroscopic description (NVE) of the system.
>
> (Rao, 1994)

While the first postulate applies to all ensembles, the second only applies to the microcanonical ensemble. The microcanonical ensemble is characterized by variables NVE.

From this postulate one can see that for the microcanonical ensemble every permissable quantum state has the same probability. Thus, one is able to determine the ensemble averages for the mechanical variables of this ensemble. With the definition of an ensemble and these two postulates it is possible to develop expressions for macroscopic thermodynamic properties in the microcanonical and other ensembles. Mechanical variables are determined by measuring the ensemble average of that variable during a simulation.

2.1.4 Definition: An Ensemble Average

For a discrete set of microstates, the ensemble average of a given thermodynamic property, $\mathcal{A}$, is:

$$\langle \mathcal{A} \rangle_{ens} = \sum_{i} \mathcal{A}_i \wp_i \tag{2.2}$$

where $\mathcal{A}_i$ is the value of the thermodynamic property for the i^{th}-ensemble state and $\wp_i$ is the probability that this i^{th}-ensemble state will occur. While Equation 2.2 is the correct discrete form, it is typically not directly used to determine the ensemble averages. Consider a set of states, i^*, which have been biased using the probability density, $\wp$, then the following equation may be used:

$$\langle \mathcal{A} \rangle_{ens} = \frac{1}{s} \sum_{i=1}^{s} A_{i^*} \tag{2.3}$$

Where s is the number of states sampled.

Clearly the determination of the probability distribution, $\wp$ is critical to determining the ensemble average. Before this is possible the partition functions are required and are the focus of the following section. Interestingly, one can see that if the probability of a given state is very low then the contribution to the average will also be low. This fact is used extensively in Monte Carlo simulations and will be discussed at length in Chapter 3.

2.2 Standard Statistical Ensembles

This section is only a brief review of the statistical mechanics of the four standard ensembles; detailed derivations may be found in Rao (1994). The four ensembles presented are the microcanonical (constant-NVE), canonical (constant-NVT), isothermal-isobaric (constant-NPT) and grand-canonical (constant-μVT). There are a number of possible

ensembles, however, only a few are of practical interest. How an ensemble is defined depends on the thermodynamic state of the system we wish to represent. The thermodynamic quantities not specified must be determined by ensemble averaging. It is instructive to note the pairing of extensive and intensive variables, shown in Table 2.1 (Kofke, 2003). Any thermodynamic state requires at least one variable from each of the

Extensive variable	Intensive variable
Internal energy, E	Temperature, T
Volume, V	Pressure, P
Number of molecules, N	Chemical potential, μ

Table 2.1: Pairing of extensive and intensive variables.

pairs of Table 2.1 to be defined. At least one extensive variable must be given in order for the absolute size of the system to be defined. Not defining the absolute system-size would result in the Gibbs-Duhem phase rule being violated. This means that an ensemble such as μPT, while conceivable, would not be practical.

2.2.1 Microcanonical Ensemble

The probability density for the microcanonical ensemble follows directly from the second postulate and is proportional to

$$\wp_{NVE} \propto \delta(\mathcal{H} - E) \tag{2.4}$$

where $\mathcal{H}$ is the Hamiltonian. The Hamiltonian is function of the phase-space $\Gamma = \mathbf{q}, \mathbf{p}$ (coordinates and momenta of the particles) and it expresses the total energy of an isolated system. The δ function simply selects all systems of a given number of molecules, N and volume, V with the energy, E. From this the partition function for the microcanonical ensemble may be written as:

$$\Omega = \sum_{\Gamma} \delta(\mathcal{H} - E) \tag{2.5}$$

The partition function represents the normalized sum over all microstates of an ensemble. In semi-classical form Equation 2.5 for an atomic system of N indistinguishable particles becomes:

$$\Omega = \frac{1}{N! h^{3N}} \int \delta(\mathcal{H} - E)\, \mathbf{dqdp} \tag{2.6}$$

Where h is Planck's constant and the integral is over 6N phase-space. A bridge equation is used to relate thermodynamic potentials from classical thermodynamics to these partition functions. Entropy, S, is the thermodynamic potential related to the microcanonical partition function, Ω, by:

$$S = k_B \ln(\Omega) \tag{2.7}$$

where k_B is the Boltzmann's constant.

2.2.2 Canonical ensemble

The probability density for the canonical ensemble is proportional to:

$$\wp_{NVT} \propto \exp(-\beta\mathcal{H}) \tag{2.8}$$

where $\beta = k_B T$ and the partition function is:

$$Q = \sum_{\Gamma} \exp(-\beta\mathcal{H}) \tag{2.9}$$

Which, in semi-classical form becomes:

$$Q = \frac{1}{N! h^{3N}} \int \exp(-\beta\mathcal{H})\, \mathbf{dqdp} \tag{2.10}$$

The appropriate thermodynamic potential related to the canonical ensemble partition function is the Helmholtz free energy, A, and is related by:

$$\beta A = -\ln(Q) \tag{2.11}$$

2.2.3 Isothermal-isobaric Ensemble

The probability density for the isothermal-isobaric ensemble is proportional to:

$$\wp_{NPT} \propto \exp[-\beta(\mathcal{H} + PV)] \tag{2.12}$$

and the partition function is:

$$\Delta = \sum_{\Gamma} \sum_{V} \exp[-\beta(\mathcal{H} + PV)] \tag{2.13}$$

Which, in semi-classical form becomes:

$$\Delta = \frac{1}{N! h^{3N}} \frac{1}{V_0} \int \left[\int \exp[-\beta(\mathcal{H} + PV)]\, \mathbf{dqdp} \right] dV \tag{2.14}$$

Where V_0 is the basic unit of volume used to keep Δ dimensionless. The appropriate thermodynamic potential related to the isothermal-isobaric ensemble partition function is the Gibbs free energy, G, and is related by:

$$\beta G = -\ln(\Delta) \tag{2.15}$$

2.2.4 Grand-canonical Ensemble

The probability density for the grand-canonical ensemble is proportional to:

$$\wp_{\mu VT} \propto \exp[-\beta(\mathcal{H} - \mu N)] \tag{2.16}$$

and the partition function is:

$$\Xi = \sum_{\Gamma} \sum_{N} \exp[-\beta(\mathcal{H} - \mu N)] \tag{2.17}$$

Which, in semi-classical form becomes:

$$\Xi = \sum_{N} \frac{1}{N! h^{3N}} \int \exp[-\beta(\mathcal{H} - \mu N)] \, d\mathbf{q}d\mathbf{p} \tag{2.18}$$

Where the sum N remains a summation because N is not continuous. The appropriate thermodynamic potential related to the grand-canonical ensemble partition function is the Hill potential, and is related by:

$$\beta PV = -\ln(\Xi) \tag{2.19}$$

2.3 Thermodynamic Averages

In this section we will focus on the mathematical forms of the averages used to relate microscopic behaviour to macroscopic behaviour.

2.3.1 Probability Distribution of Microstates

Before one can discuss thermodynamic averages, the probability distribution, $\wp$ from Equation 2.2, must be considered. While they have not been derived it should be noted that the distributions follow a certain amount of common sense. Table 2.2 shows the probabilities for the common ensembles. The probability distributions presented here can be considered as the normalized number of microstate occurrences divided by the normalized sum of states (the partition function). It is now possible to determine the thermodynamic averages. The rest of this section focuses on the derivation of the thermodynamic averages used in this dissertation.

Ensemble	All states of:	Probability distribution
Microcanonical	given NVE	$\wp_{NVE} = \delta(\mathcal{H} - E)/\Omega$
Canonical	given NVT	$\wp_{NVT} = \exp(-\beta\mathcal{H})/Q$
Isothermal-isobaric	given NPT	$\wp_{NPT} = \exp[-\beta(\mathcal{H} + PV)]/\Delta$
Grand-canonical	given μVT	$\wp_{\mu VT} = \exp[-\beta(\mathcal{H} - \mu N)]/\Xi$

Table 2.2: Probability distributions for common ensembles.

2.3.2 Internal Energy

It is possible to express the internal energy in two main parts: the kinetic-energy contribution which is a function of momenta only and the potential-energy contribution which is a function of the coordinates only.

$$E = \mathcal{H}(\mathbf{q}, \mathbf{p}) = \mathcal{K}(\mathbf{p}) + \mathcal{U}(\mathbf{q}) \tag{2.20}$$

Where $\mathcal{K}$ is the kinetic energy and $\mathcal{U}$ is the potential energy. Splitting the Hamiltonian in this manner allows one to simplify the probability distribution and partition functions expressions presented above. These expressions may be factorized into kinetic and potential parts. Consider the semi-classical form of the canonical partition function (Equation 2.10), it may be written as follows:

$$Q = \frac{1}{N!h^{3N}} \int \exp(\beta\mathcal{K})\, d\mathbf{p} \int \exp(\beta\mathcal{U})\, d\mathbf{q} \tag{2.21}$$

Now, it has been shown in Allen and Tildesley (1987) that $2\langle\mathcal{K}\rangle = 3Nk_BT$. Since the determination of the kinetic contribution is trivial, and constant for the isothermal ensembles, the focus generally shifts to the determination of the potential energy. The importance of determining the potential energy accurately and efficiently cannot be overstated, it is a corner-stone in Monte Carlo simulation.

2.3.3 Pressure

A usable definition for the determination of the pressure is not trivial. Due to the use of periodic boundaries (see Chapter 5) we are unable to simply measure the momentum flux against the boundaries of the simulation volume. There are two methods available for determining the system pressure, the first is derived from the virial theorem and the other from the thermodynamic definition of the pressure. The virial pressure is used

exclusively for the determination of pressure in systems where the potential energy of the system is volume independent (i.e. only van der Waals interactions considered). The thermodynamic pressure is used when the potential energy does depend on the volume (i.e. long-range Coulomb interactions handled using Ewald summations - Chapter 5)

Virial Pressure

Derivation of the virial pressure is presented in a number of texts, but Allen and Tildesley (1987) gives an appropriate derivation. It begins with the definition of the virial theorem:

$$\left\langle q_k \frac{\partial \mathcal{H}}{\partial q_k} \right\rangle = k_B T \tag{2.22}$$

where q_k is the generalized momentum. From this, one is able to derive an expression for the pressure in terms of an ideal contribution and a contribution due to the intermolecular forces:

$$P_{vir} V = N k_B T + \langle \mathcal{W} \rangle \tag{2.23}$$

where $\mathcal{W}$ is the internal virial. Rewriting Equation 2.23 in the more usable, instantaneous form, one has:

$$\mathcal{P}_{vir} = \rho k_B T + \frac{\mathcal{W}}{V} \tag{2.24}$$

where $\mathcal{P}_{vir}$ is the instantaneous pressure and ρ the instantaneous density of a particular microstate. It is now convenient to define the internal virial, $\mathcal{W}$. Considering a 3-dimensional system:

$$\mathcal{W} = \frac{1}{3} \sum_{i=1}^{N} \mathbf{r_i} \cdot \mathbf{f_i} \tag{2.25}$$

$$= \frac{1}{3} \sum_{i}^{N-1} \sum_{j>i}^{N} \mathbf{r_{ij}} \cdot \mathbf{f_{ij}} \tag{2.26}$$

$$= -\frac{1}{3} \sum_{i}^{N-1} \sum_{j>i}^{N} w(r_{ij}) \tag{2.27}$$

where $\mathbf{f_{ij}}$ is the force that molecule j exerts on molecule i and $\mathbf{r_{ij}} \equiv \mathbf{r_i} - \mathbf{r_j}$ is the vector between molecule centers. Hence, the intermolecular pair virial function $w(r_{ij})$ is:

$$w(r_{ij}) = r_{ij} \frac{d\mathcal{U}(r_{ij})}{dr_{ij}} \tag{2.28}$$

where $\mathcal{U}(r_{ij})$ is the pair-wise intermolecular potential and r_{ij} is the intermolecular separation between the pair. Thus, one may write the complete expression for the virial

pressure as:

$$P_{vir} = \left\langle \rho k_B T - \frac{1}{3V} \sum_{i=1}^{N} \sum_{j>i} r_{ij} \frac{d\mathcal{U}(r_{ij})}{dr_{ij}} \right\rangle \tag{2.29}$$

Thus, the above expression is used when the potential does not explicitly depend on the volume. If it does, however, then one must use the thermodynamic pressure.

Thermodynamic Pressure

Using the method described in Hummer and Grønbech-Jensen (1998), the thermodynamic pressure is derived from the following relationship:

$$P_{therm} = -\left(\frac{\partial A}{\partial V}\right)_{T,N} \tag{2.30}$$

and it can be shown that this can also be expressed in terms of an ideal contribution and a contribution due to intermolecular forces:

$$P_{therm} = \rho k_B T - \left\langle \frac{\partial \mathcal{U}}{\partial V} \right\rangle \tag{2.31}$$

One can see that if the potential energy, $\mathcal{U}$, does not depend explicitly on the volume then:

$$\frac{\partial \mathcal{U}}{\partial V} = \sum_i \frac{\partial \mathcal{U}}{\partial \mathbf{r}_i} \cdot \frac{\partial \mathbf{r}_i}{\partial V} = \frac{1}{3V} \sum_i \frac{\partial \mathcal{U}}{\partial \mathbf{r}_i} \cdot \mathbf{r}_i \tag{2.32}$$

which results in the same expression as for the virial pressure. Equation 2.31 will be used in Simulation Techniques (Section 5.4) to derive an expression for the pressure when using the Ewald summation.

2.3.4 Excess Chemical Potential

Mechanical properties such as $< N >$, $< V >$, and other properties which can be determined from these, are easily determined from a simulation. Unfortunately, the determination of thermal properties are not straightforward since these properties depend on the total volume of phase-space (Siepmann, 1990). An attempt to measure them would simply lead to a very poor estimate of it. Thus, special techniques are required. The most common method (and one of the earliest introduced) of determining the chemical potential is the 'particle-insertion method' (Widom, 1963; Widom, 1982). The ratio of two partition functions Q_{N+1}/Q_N are taken:

$$Q_{N+1} = Q_N V \left\langle \exp\left(-\frac{\mathcal{U}_{test}}{k_B T}\right) \right\rangle \tag{2.33}$$

where U_{test} is the potential energy of the $(N+1)^{th}$ particle. Now from the definition of the chemical potential (Siepmann, 1990):

$$\mu = \left(\frac{\partial A}{\partial N}\right)_{V,T} = \lim_{N,V\to\infty}\left[-k_B T \ln\left(\frac{Q_{N+1}}{\Lambda^3 N Q_N}\right)\right] \tag{2.34}$$

where Λ is the de Broglie wavelength. Thus, for the excess chemical potential one is able to write:

$$\mu_{ex} \equiv \mu - \mu_{id} \tag{2.35}$$

$$= -k_B T \ln\left(\frac{Q_{N+1}}{Q_N V}\right) \tag{2.36}$$

$$= -k_B T \ln\left\langle\frac{\mathcal{U}_{test}}{k_B T}\right\rangle \tag{2.37}$$

In this way the average probability of acceptance of a Monte Carlo move consisting of the addition of this 'ghost' molecule is used to determine the excess chemical potential. Clearly, the addition of this 'ghost' molecule must not actually affect the system. The above derivation is for the canonical ensemble. The expression for the excess chemical potential in the grand-canonical ensemble is the same. For the microcanonical and isothermal-isobaric ensembles the expressions are slightly different; these can be found in Allen and Tildesley (1987).

The Widom insertion method is adequate for low-density monatomic molecules. Clearly, the hydrocarbons dealt with in this study do not fall into this category. It can be shown (Frenkel et al., 1991; de Pablo et al., 1992*a*; Frenkel and Smit, 2002) that the ratio of the partition functions can be related to the Rosenbluth weight. Thus, the above expression (Equation 2.34) for the excess chemical potential may be written in terms of the exponential of the Rosenbluth weight, W as:

$$\mu_{ex} = -k_B T \ln\langle W\rangle \tag{2.38}$$

Using this definition one is able to measure the excess chemical potential during the simulation of long-chain molecules. The Rosenbluth weight will be defined and discussed in Simulation Techniques (Chapter 5).

2.4 Radial Distribution Function

The radial distribution function (RDF) is a very useful quantity. It is a pair correlation function which describes how, on average, the atoms in a system are radially packed

around each other (*Democritus - Radial Distribution Function*, 2004). The RDF is an effective method of characterizing the structure of simulated molecular systems. RDFs may be determined experimentally, thus one is able to compare predicted fluid conformations with actual measured conformations. The RDF is calculated measuring the density of particles at a distance r from the reference particle and comparing this with the bulk density of the system. Figure 2.1 shows a system split into six regions, at which Equation 2.39 would be used. While Figure 2.1 shows graphically (for a 2D system) how the

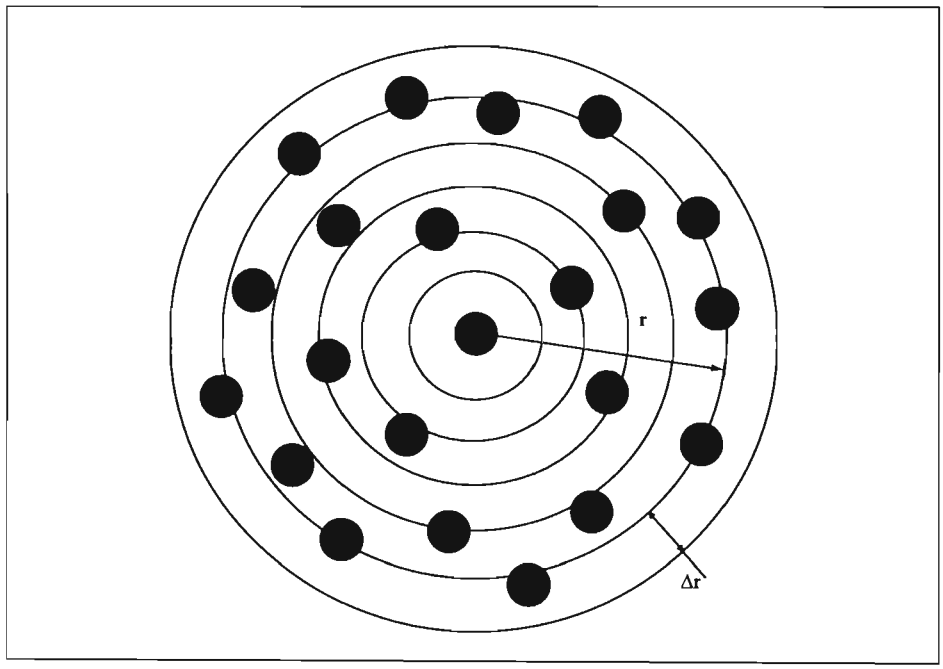

Figure 2.1: Graphical representation of the RDF calculation

RDF is calculated, mathematically, it may be determined using the following expression:

$$\mathbf{g(r)} = \frac{\mathbf{n(r)}}{(4\pi\rho\mathbf{r}^2\Delta\mathbf{r})} = \frac{\rho(\mathbf{r})}{\rho} \tag{2.39}$$

where $\mathbf{g(r)}$ is the RDF, $\mathbf{n(r)}$ is the mean number of atoms within a shell of width $\Delta\mathbf{r}$ at a distance $\mathbf{r}$, $\rho(\mathbf{r})$ is the radial density and ρ is the bulk or mean system density.

Characterization is not the only use for the RDF. It is possible to write the ensemble average of a property $\mathcal{A}$ as (Allen and Tildesley, 1987):

$$\langle\mathcal{A}\rangle = \frac{1}{2}N\rho\int_0^\infty \mathcal{A}(r)g(r)4\pi r^2\,dr \tag{2.40}$$

Which may be used to determine the system energy, density and chemical potentials. This fact is used in the Simulation Techniques Chapter (Chapter 5) to account for tail corrections.

Chapter 3

Monte Carlo Simulation

3.1 The Monte Carlo Method

It may appear that the most simple and direct method for determining the thermodynamic averages would be to analytically determine the partition functions shown in Table 2.2. Unfortunately, analytical solutions of the partition functions are not possible for any but the most simple of systems (in terms of the number of molecules and dimensions). This difficulty was discovered early on in the history of molecular simulation and thus the Monte Carlo (MC) method was introduced (Metropolis et al., 1953). Allen and Tildesley (1987) and Frenkel and Smit (2002) discuss in depth the many advantages of the MC method.

3.1.1 Random Sampling Method

The the most basic type of MC method is the random sampling method. Consider a function, $f(x)$, integrated over a range a to b:

$$I = \int_a^b f(x)dx \tag{3.1}$$

This can be rewritten as the following expression:

$$I = (b - a)\langle f(x)\rangle \tag{3.2}$$

Where $\langle f(x)\rangle$ is the unweighted average of $f(x)$ over $[a, b]$. Equation 3.2 may now be used to solve the integral. This is done by determining $\langle f(x)\rangle$ using a large set of randomly distributed values of x on $[a, b]$. As the number of values used to determine the average increases, so Equation 3.2 will yield better results for I.

This method is conceptually simple, but it is not directly applicable to the partition functions and property averages discussed in Chapter 2. The reason is that many of the randomly selected configurations (the x's) would not contribute significantly to the averages. A possible solution to this difficulty would be to sample from a nonuniform distribution over the integration range and then correct for it. Frenkel and Smit (2002) show how using a probability density, $\wp(x)$, to sample points on nonuniform distribution may be used to improve the the prediction of the integral. Using this probability density one may write Equation 3.1 as:

$$I = \int_a^b \frac{f(x)}{\wp(x)}\wp(x)dx \tag{3.3}$$

$$= \int_a^b \frac{f[u(x)]}{\wp[x(u)]}du \tag{3.4}$$

$$\approx \frac{(a-b)}{\tau}\sum_{i=1}^{\tau}\frac{f[x(u_i)]}{\wp[x(u_i)]} \tag{3.5}$$

Where τ is the number of random samples taken from the nonuniform distribution. To use Equation 3.5 the probability density $\wp(x)$ is required. Unfortunately, the probability distributions given in Chapter 2 are not known *a priori* since that would require knowing the partition function *a priori* (Table 2.2). It is possible however to know the probability distributions relative to another state though, and thus a slightly different method must be used.

3.1.2 Metropolis Method

The Metropolis method involves the construction of an importance-weighted random walk through phase-space, where the distributions are nonnegligible. Frenkel and Smit (2002) compare this method with attempting to determine the average depth of the Nile by only taking measurements within the Nile, whereas using the random sampling method is sampling over the whole of Africa to determine the same average. In this way one is able to determine the equilibrium averages if one can 'get into' equilibrium phase-space and determine the thermodynamic averages, but is still not able to directly determine the partition function.

Metropolis et al. (1953) originally developed this method for a canonical (NVT) ensemble with only changes in the particles positions within the box (displacement moves). The canonical ensemble will be discussed later in the chapter. The remainder of this section will be devoted to deriving the Metropolis scheme in such a way that one may use

the expressions to develop and validate acceptance criteria for new or old trial moves.

The first step is to determine the probability distribution, $\wp$, for the ensemble of interest. Since the distributions are considered relatively, the change in the state from an old configuration, o, to a new configuration, n will be used. In order to relate these two states one constructs the biased-walk as a Markov chain. A Markov chain is a sequence of trials that satisfies (Allen and Tildesley, 1987):

- There are a finite number of outcomes for each trial, and
- The outcome of each trial depends only on the trial that immediately precedes it.

Thus, the change between any two states (say old to new) are linked by a transition probability matrix, $\pi(o \rightarrow n)$. Clearly these probabilities must maintain the equilibrium conditions once it has been reached, thus the condition of detailed balance is imposed. This means that the probability of change in state from the old to the new is equal to the probability of change from the new to the old states:

$$K(o \rightarrow n) = K(n \rightarrow o) \tag{3.6}$$

While detailed balance is sufficient but not necessary, Allen and Tildesley (1987) and Frenkel and Smit (2002) suggest its use. Thus, from Equation 3.6 one may write:

$$\wp(o) \times \pi(o \rightarrow n) = \wp(n) \times \pi(n \rightarrow o) \tag{3.7}$$

It is possible to consider the transition matrix, $\pi(o \rightarrow n)$, in terms of a probability of generating a new state, $\alpha(o \rightarrow n)$*, and the probability of accepting this generated state:

$$\pi(o \rightarrow n) = \alpha(o \rightarrow n) \times acc(o \rightarrow n) \tag{3.8}$$

So one may write Equation 3.7 as:

$$\wp(o) \times \alpha(o \rightarrow n) \times acc(o \rightarrow n) = \wp(n) \times \alpha(n \rightarrow o) \times acc(n \rightarrow o) \tag{3.9}$$

In the original Metropolis scheme $\alpha(o \rightarrow n)$ was assumed to be symmetrical (i.e. the forward change is as likely to happen as the reverse change):

$$\alpha(o \rightarrow n) = \alpha(n \rightarrow o) \tag{3.10}$$

*Often called the underlying matrix of the Markov chain.

Thus, one is able to cancel out the α-terms from Equation 3.9 and rewrite it in a more appropriate form:

$$\frac{acc(o \rightarrow n)}{acc(n \rightarrow o)} = \frac{\wp(n)}{\wp(o)} \tag{3.11}$$

Now, since $acc(n \rightarrow o) \leq 1$ one may write:

$$acc(o \rightarrow n) = \begin{cases} \frac{\wp(n)}{\wp(o)} & \text{if } \wp(n) < \wp(o) \\ 1 & \text{if } \wp(n) \geq \wp(o) \end{cases} \tag{3.12}$$

Finally one is able to write the general expression for the acceptance criteria for a trial move as:

$$acc(o \rightarrow n) = min\left(1, \frac{\wp(n)}{\wp(o)}\right) \tag{3.13}$$

In practice, a randomly generated number (between 0 and 1) is compared to the ratio $\wp(n)/\wp(o)$ and if the random number is larger than the ratio then the attempted move is accepted. One can see that if the ratio is larger than 1 it is always accepted. If the ratio is larger than 1 then this generally means that there has been a decrease in the systems potential energy, and since a system is at its minimum energy at equilibrium, one can see that this scheme will automatically force the system into equilibrium phase-space. If an attempted trial move is not accepted then the original state must be recounted. This is because (Frenkel and Smit, 2002):

$$\pi(o \rightarrow o) = 1 - \sum_{n \neq o} \pi(o \rightarrow n) \tag{3.14}$$

Thus, when a trial move is not accepted then $\pi(o \rightarrow o) > 0$ and the old state must be re-counted.

3.1.3 Trial Moves

There are three basic categories for trial moves used in MC simulation today:

- Displacement moves, where the molecules' positions or conformations are changed;
- Volume changes (e.g. NPT);
- Molecule insertions and deletions (e.g. μVT).

The particular trial moves used in a simulation depends on the ensemble. Obviously, in the canonical ensemble the volume change and molecule insertion moves have no meaning. To date many trial moves have been developed, but most of them fall into one of

the above categories. For example, Chen and Siepmann (2000) introduces an aggregated volume move, which attempts to create and break strong intermolecular bonds (such as hydrogen bonding), it is still a displacement type move.

When it comes to developing new or alternative trial moves one must be very careful to maintain, at the very least, the "balance condition" but it is preferable to maintain the "detailed-balance condition" (Frenkel and Smit, 2002). This must be done to ensure that no biases are introduced into the scheme. Each of the trial moves, displacement, volume change, and molecule insertions, will be developed and discussed within their appropriate ensembles. During this section the term molecules has been used. In most cases the trial moves are developed in terms of particles and then extended to molecules using Configurational-bias Monte Carlo (see Chapter 5). Thus, for the rest of this chapter, the systems considered will be monatomic in nature.

3.2 Monte Carlo Simulations in Classical Ensembles

Having looked at the basic concept of the Monte Carlo simulation method, it is now appropriate to discuss its application to three important classical ensembles. These ensembles, the canonical, isothermal-isobaric and grand-canonical, form the basis of most of the Monte Carlo simulations done today. It is these three ensembles that are combined to produce the Gibbs Ensemble method (Chapter 6). In this section the basic ensemble schemes and acceptance criteria will be developed from the relevant probability distributions. As mentioned above, the condition of detailed balance is too strong a condition, but in using it one is guaranteed to have correct sampling.

3.2.1 Canonical Ensemble

The canonical (or NVT) ensemble is the ensemble originally used by Metropolis et al. (1953). In this ensemble the number of molecules, N, the volume, V, and the temperature, T, are held constant. This limits the possible types of moves used in this ensemble to displacement moves (see Figure 3.1). The probability distribution (Equation 2.8) may be written in terms of the potential energy as following:

$$\wp_{N,V,T}(\mathbf{r}^N) \propto \exp[-\beta \mathcal{U}(\mathbf{r}^N)] \tag{3.15}$$

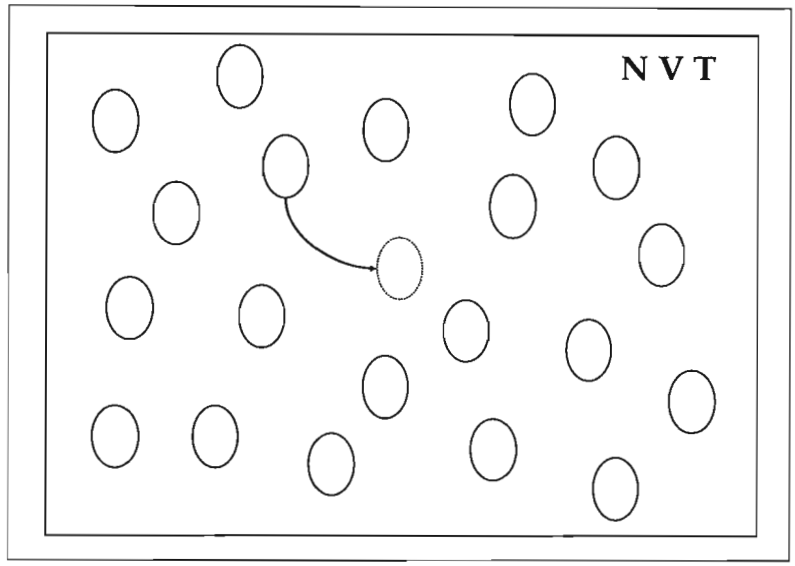

Figure 3.1: The canonical (NVT) ensemble. Only displacement type trial moves are possible in this ensemble.

From this distribution one can see that only displacement type moves are possible in this ensemble. The following MC scheme may be used:

1. Select a particle at random and determine $\mathcal{U}(o)$.
2. Randomly displace the particle (as shown in Figure 3.1) using:

$$\mathbf{r}(n) = \mathbf{r}(o) + \Delta\mathbf{r}_{max}(Rnd - 0.5), \tag{3.16}$$

   where $\mathbf{r}$ is vector coordinate position of the particle, Rnd is a random number $[0, 1]$ and $\Delta\mathbf{r}_{max}/2$ is the maximum displacement.
3. Calculate the new state's potential energy, $\mathcal{U}(n)$.
4. The acceptance probability is then obtained by substituting Equation 3.15 into Equation 3.13:

$$acc(o \rightarrow n) = min\,(1, exp\{-\beta[\mathcal{U}(n) - \mathcal{U}(o)]\}) \tag{3.17}$$

It is important to note the exact form of Equation 3.16. Since Rnd varies between 0 and 1 the change in position varies between $-\Delta\mathbf{r}_{max}/2$ and $+\Delta\mathbf{r}_{max}/2$. So one can see that the reverse trial move is as possible as the forward move is, thus maintaining symmetry in the underlying Markov matrix (α). Now say the following was used to change the position rather:

$$\mathbf{r}(n) = \mathbf{r}(o) + Rnd \times \Delta\mathbf{r}_{max} \tag{3.18}$$

While at first glance this appears valid, further investigation shows that this expression displaces the particle on $[0, \Delta \mathbf{r}_{max}]$. Clearly the reverse move is not possible, and thus the underlying Markov matrix (α) is not symmetrical.

Another aspect of importance is how large one should make $\Delta \mathbf{r}_{max}$. One can see that as $\Delta \mathbf{r}_{max}$ is made larger greater phase-space will be explored but the number of accepted moves will rapidly decrease. And as one reduces the size so the number of accepted moves will increase, but the phase-space explored will decrease. There have been a number investigations into this, but the optimum size varies from simulation to simulation. Generally, $\Delta \mathbf{r}_{max}$ is modified during the running of the simulation to maintain a particular acceptance ratio. Again, there is much debate as to the optimal acceptance ratio but studies have suggested values ranging from about 20% for hard-core systems up to about 50%.

3.2.2 Isobaric-Isothermal Ensemble

The isothermal-isobaric ensemble is possibly the most common ensemble used in Monte Carlo simulation after the Gibbs ensemble method. Of all the ensembles it is the closest to actual experiment in that both the temperature and pressure are held constant. Two trial moves are possible in this ensemble, displacement moves (as in the NVT ensemble) and volume change moves (see Figure 3.2). The probability distribution for the NPT

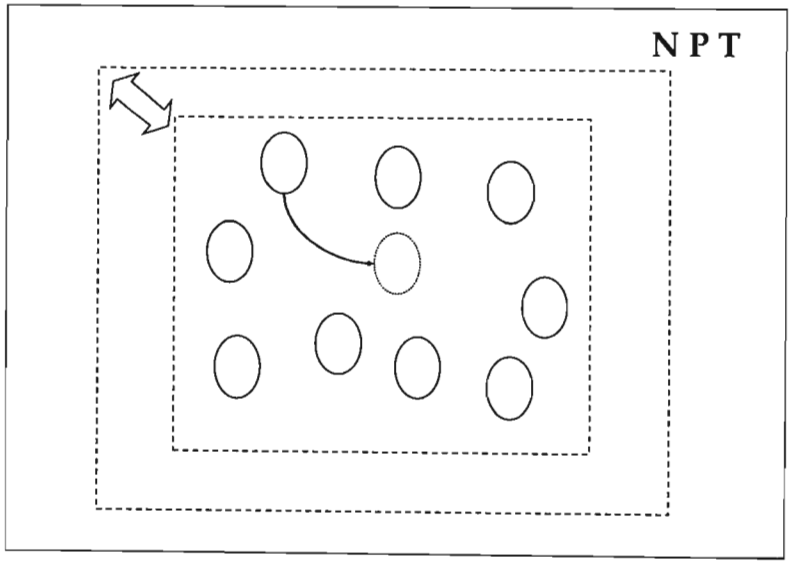

Figure 3.2: The isothermal-isobaric (NPT) ensemble. Here displacement and volume change type trial moves are possible.

ensemble, as it is shown in Equation 2.12, is not convenient. Rather, the coordinates are written in terms of scaled coordinates:

$$\mathbf{s}^N = \mathbf{r}^N / L \tag{3.19}$$

Where $L = V^{1/3}$ is the box length. Thus, an additional factor of V^N must be considered. The probability distribution with scaled coordinates is:

$$\wp_{N,P,T}(\mathbf{s}^N, V) \propto \exp[-\beta \mathcal{U}(\mathbf{s}^N, V) + PV + N \ln V] \tag{3.20}$$

A similar MC scheme as that used for the canonical ensemble may be used here:

1. Determine current system energy $\mathcal{U}(o)$.
2. Randomly select between the possible trial moves:
   (a) Displacement of a particle,
   (b) Change in the box volume.
3. If a particle displacement move is attempted then use Equation 3.16, else the simulation box volume (see Figure 3.2) is changed using:

$$V(n) = V(o) + \Delta V_{max}(Rnd - 0.5), \tag{3.21}$$

   where the maximum change in volume is $\Delta V_{max}/2$
4. Calculate the new states potential energy $\mathcal{U}(n)$.
5. The acceptance probability is then obtained by substituting Equation 3.20 into Equation 3.13:

$$\begin{aligned} acc(o \to n) &= \min(1, \exp\{-\beta[\mathcal{U}(n) - \mathcal{U}(o)] \\ &\quad + P(V(n) - V(o)) - N \ln(V(n)/V(o))\}) \end{aligned} \tag{3.22}$$

From Equation 3.22 one can see that if a displacement move is attempted then the acceptance criteria reduces to the same expression as for the NVT ensemble. In some cases, instead of a random walk on V, one might choose a random walk on say $\ln V$ instead. In this case the probability distribution must be modified and thus the acceptance criteria will be different. Frenkel and Smit (2002) show how these must be altered for different walks on V.

Generally one has to recalculate all the interactions when the volume is changed. For certain types of potential energy functions, where they are written as a sum of powers of the interaction distances (Lennard-Jones 12-6 and Coulomb interactions), if scaled coordinates are used then:

$$\begin{aligned} \mathcal{U}_n &= \sum_{i<j} \epsilon \left(\frac{\sigma}{r_{ij}}\right)^n = \sum_{i<j} \epsilon \left(\frac{\sigma}{L s_{ij}}\right)^n \\ &= \left(\frac{1}{L}\right)^n \sum_{i<j} \epsilon \left(\frac{\sigma}{s_{ij}}\right)^n \end{aligned} \tag{3.23}$$

Thus, when the volume is changed, the new states potential is very easy and quick to determine:

$$\mathcal{U}_n(L') = \left(\frac{L}{L'}\right)^n \mathcal{U}_n(L) \tag{3.24}$$

It is important to remember all coordinate related terms (such as cutoff radii) are scaled as well. In the extension to molecules it is important to note that only the center-of-mass positions are scaled when the volume is changed. This means that the molecule must be rigid. Unfortunately, this fact rules out its use for hydrocarbons since they are deformable molecules.

3.2.3 Grand-Canonical Ensemble

The grand-canonical ensemble is particularly different from the previous two ensembles in that it holds the chemical potential constant instead of the number of particles. Along with chemical potential, the volume and temperature are also kept constant. Here the only extensive variable is the volume. This ensemble is best understood as a volume, V, surrounded by an ideal gas composed of the same particles as within the volume at a chemical potential μ. There are particles entering and leaving this volume in such a way as to maintain the chemical potential in the volume. Only the particles within the volume interact with other particles. As with the isothermal-isobaric ensemble it is convenient to work with scaled coordinates. Thus, one may write the probability distribution as:

$$\wp_{\mu,V,T}(\mathbf{s}^N, N) \propto \frac{\exp(\beta\mu N)V^N}{\Lambda^{3N}N!} \exp[-\beta\mathcal{U}(\mathbf{s}^N)] \tag{3.25}$$

A similar MC scheme as that used for the previous ensembles is used for μVT simulations:

1. Determine current system energy $\mathcal{U}(N)$.

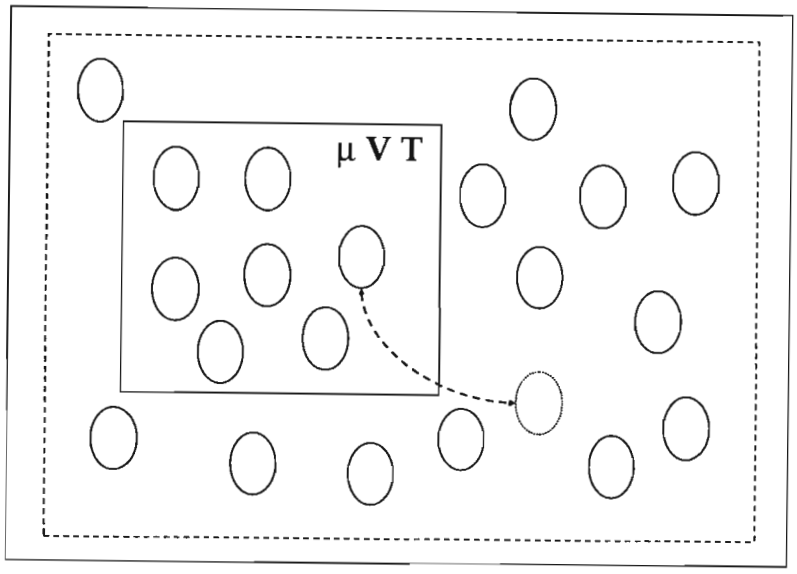

Figure 3.3: The grand-canonical μVT) ensemble. Here displacement and particle insertion type trial moves are possible.

2. Randomly select between the possible trial moves:

   (a) Displacement of a particle,

   (b) Insert or remove a particle.

3. If a particle displacement move is attempted then use Equation 3.16, else attempt a particle insertion or removal (see Figure 3.3). Whether an insertion or removal move is attempted must be selected randomly.

4. Calculate the new states potential energy: $\mathcal{U}(N+1)$ or $\mathcal{U}(N-1)$.

5. The acceptance probability for a particle insertion is given by:

$$acc(N \rightarrow N+1) = \min\left[1, \frac{V}{\Lambda^3(N+1)} \exp\{-\beta[\mathcal{U}(N+1) - \mathcal{U}(N) - \mu]\}\right] \quad (3.26)$$

   And for a particle removal:

$$acc(N \rightarrow N-1) = \min\left[1, \frac{\Lambda^3 N}{V} \exp\{-\beta[\mathcal{U}(N-1) - \mathcal{U}(N) + \mu]\}\right] \quad (3.27)$$

Since this ensemble inserts particles one can see that as the density of the system simulated increases so the difficulties in inserting particles increases. This problem is magnified for molecules and particularly for long-chain molecules. Methods such as CBMC (Chapter 5) may be used but these method do not eliminate the difficulties associated

with this ensemble, only delay them until higher densities or longer molecules are attempted. This ensemble is the basis of the Histogram Reweighting method which used for predicting phase equilibria (Chapter 6).

Chapter 4

Calculation of Potential Energy

Molecular simulations have been shown to rely heavily on the determination of the potential energy of a system. The potential energy of a system is generally determined by splitting it into many parts. The two major ones being the potential due to the molecules' conformation (intra-), and the potential due to the interactions between the molecules in the system (inter-). Only pairwise additive forces (the interactions between pairs of sites on molecules) have been considered in this study. Generally the many-body effects are taken into account implicitly during the parametrization of particular potential models.

4.1 Potential Energy

Just as the internal energy is split into in two parts, kinetic and potential, so the potential energy may also be split into many parts. This fact is taken advantage of in configurational-bias Monte Carlo. The potential energy of the system can be split into two parts:

$$\mathcal{U} = u^{int} + u^{ext} \tag{4.1}$$

where u^{int} is the potential energy due the intramolecular interactions (i.e. potential due to the molecular conformation) and u^{ext} is the potential energy due to intermolecular interactions. This split is completely arbitrary and may be optimized for a particular application, although care must be taken not to split dependent interactions.

4.1.1 Intramolecular Components

In dealing with the intramolecular interactions, it is again possible to split apart the potential. Generally the internal potential is split into the following components:

$$u^{int} = u_{bond} + u_{bend} + u_{tors} + u_{nonbond} \tag{4.2}$$

where u_{bond} is the potential energy due to bonded interactions, u_{bend} is the potential due to the bending interactions, u_{tors} is the potential due to the torsional interactions and $u_{nonbond}$ is the potential due to non-bonded intramolecular interactions.

Clearly there are no intramolecular interactions when particles are considered or when a molecule is approximated as a single unit (e.g. methane with the hydrogens collapsed into central carbon atom). The molecules are often grouped so that the number of units in the pseudo-molecule is less than the actual number of particles in the molecule*. How physically representative this split is generally depends on the size of the molecule and the number of molecules used in the simulation. Grouped units, such as a methyl-group, are often referred to as pseudo-atomic units as well.

Bond stretching

The most basic component of the intramolecular interactions. The bond stretching component is considered for all molecules with two or more units. Two of the most common functional forms used to describe the bonding energies have been shown here. The first assumes that the bond length is constant:

$$u_{bond} = \begin{cases} u_{bond}(r) & r = r_{fixed} \\ \infty & r \neq r_{fixed} \end{cases} \tag{4.3}$$

where l_{fixed} is the fixed bond-length. Here one can see that any bond length longer or shorter than l_{fixed} is not considered. In practical applications there is a very small tolerance on fixed length. The other functional form assumes that the bonding energies may be described by a harmonic potential:

$$u_{bond} = \frac{k_r}{2}(r - r_0)^2 \tag{4.4}$$

*This grouping techniques is typically applied intuitively on a even smaller level, often without notice, to an atom where its constituents (protons, neutrons and electrons) are grouped into a single atom.

where k_r is a constant and r_0 is the equilibrium bond length. Typically the equilibrium bond length is set to the experimentally measured equilibrium bond length. Other functions, such as the Morse function, are also used to describe bond stretching.

Bond bending

This component only has meaning if the given molecule has three or more units. An angle θ is defined as the acute angle formed by two bonds joining a central pseudo-atom. As with bond-stretching, there are two functional forms used to describe the bond bending energies. These forms are the same as used in bond stretching, the first assumes a constant bond angle:

$$u_{bend} = \begin{cases} u_{bend}(\theta) & \theta = \theta_{fixed} \\ \infty & \theta \neq \theta_{fixed} \end{cases} \tag{4.5}$$

where θ_{fixed} is the fixed bend-angle. And the second assumes a harmonic potential:

$$u_{bend} = \frac{k_\theta}{2}(\theta - \theta_0)^2 \tag{4.6}$$

where k_θ is the bending constant and θ_0 is the equilibrium bending angle. As with the bond stretching, the equilibrium bending angle is typically set to the experimental values.

Torsions

There must be a minimum of four atomic-units before one can define this type of intramolecular motion. It is prudent to first define the dihedral angle, ϕ. The dihedral angle is the 'amount of twisting' experienced on a bond. Figure 4.1 shows this twisting around the bond between atomic-units B and C. This angle is measured between two planes, one formed from A-B-C and the other formed from B-C-D. Generally, ϕ is defined as zero when the A-B-C-D is cis (as is the case in Figure 4.1) but there are cases where the trans form is used as zero.

The motion of the dihedral angles, ϕ, is generally described by a cosine series, the following is the standard form used Jorgensen et al. (1984):

$$u_{tors} = c_0 + c_1[1 + \cos(\phi)] + c_2[1 - \cos(2\phi)] + c_3[1 + \cos(3\phi)] \tag{4.7}$$

where c_i are the torsional constants. It is possible to include the c_0 with the other c_i so the torsional potential is often stated without it. This torsional potential is used by many

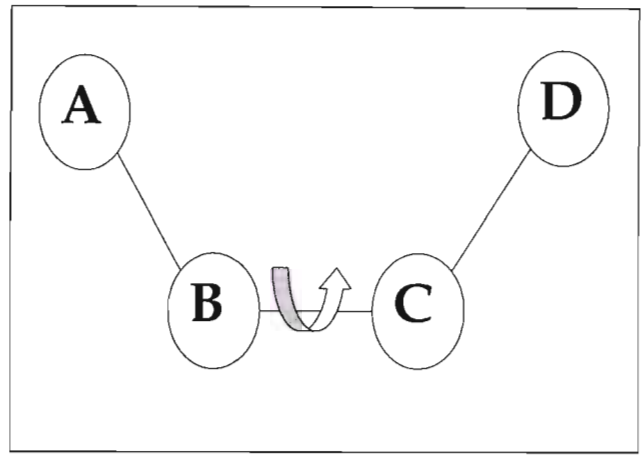

Figure 4.1: Graphical representation of the dihedral angle, ϕ.

recently parameterized potential models. Figure 4.2 (Jorgensen et al., 1984) shows how the torsional potential varies with the dihedral angle, ϕ, for the rotation about the C-C bond. It should be noted that Equation 4.7 is an even function, which can also be seen

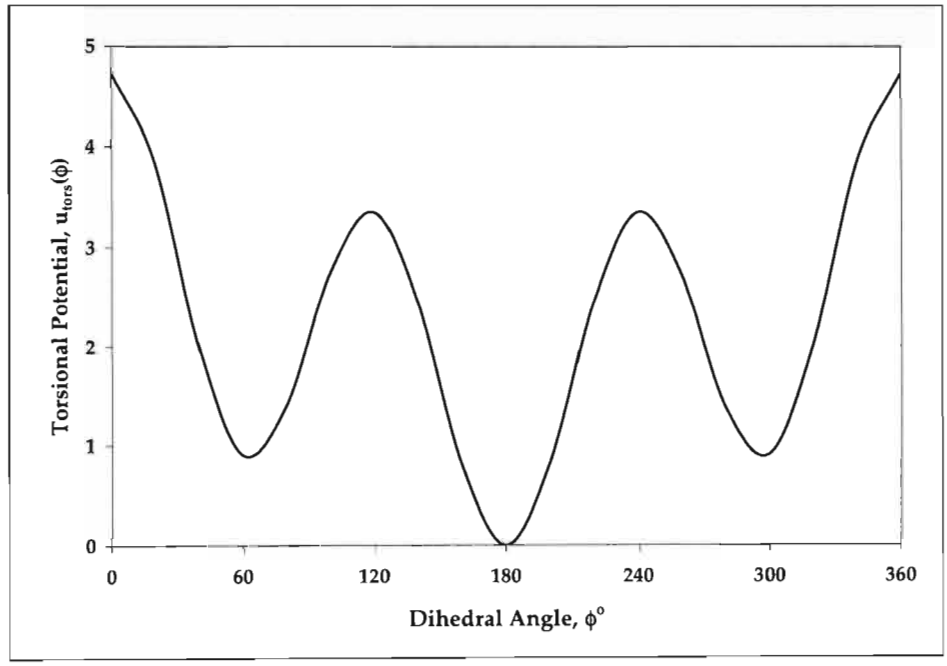

Figure 4.2: Torsional potential function for rotation about an OPLS-UA C-C bond.

in Figure 4.2. This means that the orientation at which the molecules are considered is unimportant. What is important is the definition of the dihedral angle, ϕ. Care must be taken to measure this angle correctly during simulations.

Non-bonded Intramolecular Interactions

All previously discussed intramolecular interactions have dealt with directly connected pseudo-atomic units. Now, for larger hydrocarbons, it is quite conceivable that indirectly connected parts of the same hydrocarbon may well interact. The question then becomes: at what hydrocarbon size does this begin to occur? Clearly, a three-unit hydrocarbon would not have this sort of interaction; but what about a four- or five-unit hydrocarbons? Cyclic rings are observed experimentally for five-unit hydrocarbons, thus it is very likely there is a noticeable interaction in non-cyclic hydrocarbons.

Generally this interaction is known as the '1-4 non-bond interaction' and incorporates the non-bonded interactions between two pseudo-atomic units separated by three or more bonds. Thus, for a six-unit hydrocarbon, A-B-C-D-E-F, there would be six interactions:

- $A \longleftrightarrow D, A \longleftrightarrow E, A \longleftrightarrow F$;
- $B \longleftrightarrow E, B \longleftrightarrow F$;
- $C \longleftrightarrow F$.

The non-bonded interactions considered can be either van der Waals or Coulomb interactions. Generally, in a particular simulation, the same functional descriptions used for intermolecular interactions are used here, although they may be scaled by an arbitrary factor.

4.1.2 Intermolecular Components

The external potential, which describes the intermolecular interactions, is also split into many parts. There are many different intermolecular interactions. Generally though, the van der Waals potential is used to describe the non-polar interactions. For the charged systems, Coulomb interactions ($u_{Coul} \propto r^{-1}$) or dipole interactions ($u_{Dip} \propto r^{-3}$) are used to described the forces. For this study Coulomb interactions have been used to describe the charged interactions. Thus, one can write the general split of the external potential energy as:

$$u^{ext} = u_{vdW} + u_{Coul} \tag{4.8}$$

where u_{vdW} is the potential due the van der Waals interactions and u_{Coul} the potential due to Coulomb interactions. Figure 4.3 shows the van der Waals and Coulomb potentials as a function of the intermolecular separation, r_{ij}.

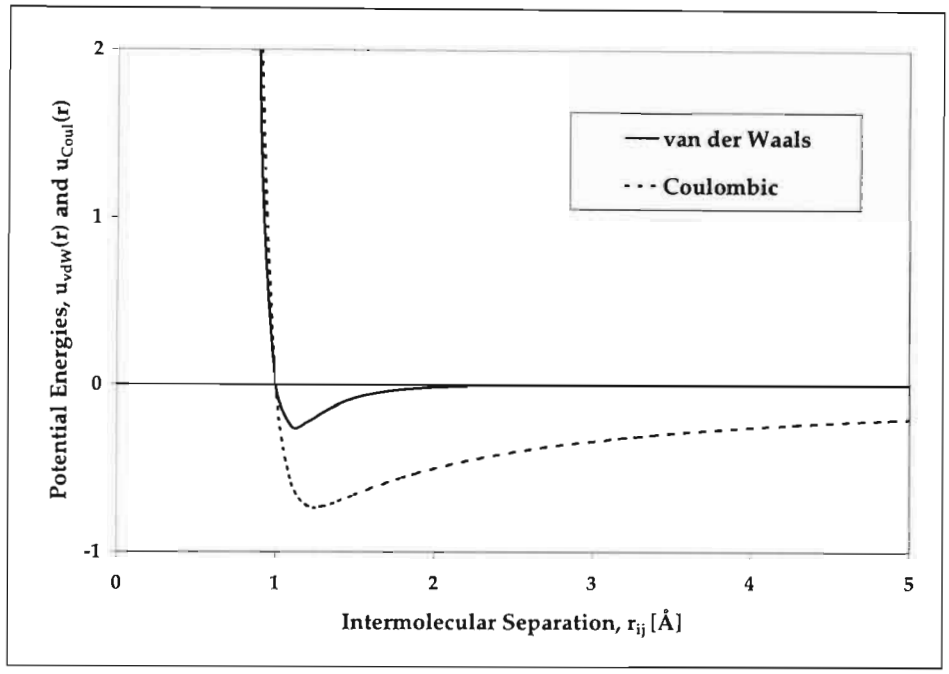

Figure 4.3: Intermolecular potentials used to described the non-bonded interactions.

van der Waals Interactions

There are many possible functional forms which have been used to describe van der Waals interactions. These have evolved over the course of the history of molecular simulation. A detailed review of the pairwise potentials used for the van der Waals interactions may be found in Sadus (1999). One of the most common functional forms used to describe the van der Waals interactions is the Lennard-Jones (LJ) 12-6 potential model:

$$u_{vdW} = u_{LJ}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{4.9}$$

where r_{ij}, ϵ_{ij} and σ_{ij} are the LJ separation, well depth and size for the pair of atoms i and j.

The attractive (or dispersive) part of the LJ potential (r^{-6}) comes from the London dispersion formula, however the repulsive part (r^{-12}) has no physical basis. The best explanation for using the power 12 is that it is double 6 (the dispersive terms power), which means that one simply needs to calculate the $(\sigma_{ij}/r_{ij})^6$-term and then square it

to determine the $(\sigma_{ij}/r_{ij})^{12}$-term. As modern computer speeds increase so the more accurate functional forms (such as the exponential function) will be used more often. It is with this in mind that potential models such as the Buckingham exponential-6 have been developed (Errington and Panagiotopoulos, 1999*a*; Errington and Panagiotopoulos, 1999*b*). This potential model has the form:

$$u_{exp-6}(r_{ij}) = \begin{cases} \frac{\epsilon}{1-\frac{6}{\omega}} \left[\frac{6}{\omega} \exp\left(\omega \left[1 - \frac{r_{ij}}{r_m}\right]\right) - \left(\frac{r_m}{r_{ij}}\right)^6 \right] & r_{ij} > r_{max} \\ \infty & r_{ij} < r_{max} \end{cases} \tag{4.10}$$

where ϵ, r_m and ω are the parameters used in this model. Parameter r_m is the radial distance at which the exponential-6 potential is a minimum and r_{max} is the smallest positive value for which $du(r)/dr = 0$. Improvements in the prediction of vapour-liquid properties cannot, however, be attributed solely to the use of the use of this exponential function since the Buckingham exponential-6 potential model also makes use of an extra parameter (ω) for every atom/group.

Coulomb Interactions

The Coulomb interactions are described using Coulomb's law of electrostatic interactions between charges:

$$u_{Coul}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{4.11}$$

where q_i and q_j are the partial charges. It can be seen that the potential decays more slowly than r^{-3}, i.e. long-ranged. The Ewald summation technique is used for dealing with these interactions and will be discussed in Simulation Techniques (Chapter 5).

Mixing Rules

In most simulation situations there are many different pseudo-atomic units. Since different units generally have different LJ parameters, there is a need to determine how these particle parameters are combined. All the potential models used in this study, use the same combining rules. They are known as the Lorenz-Berthelot combining rules:

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \tag{4.12}$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \tag{4.13}$$

These are not the only mixing rules used and they have no physical basis but are the most commonly used.

4.2 Potential Models Investigated

Potential models (the term 'force fields' is commonly used as well) are used in molecular simulations to calculate the potential energy. Hydrocarbons represent relatively simple molecular structures but are still of great industrial interest, thus there is quite a large amount of experimental data for the shorter hydrocarbons. For these very reasons, a large effort has been put into the development of transferable force fields for hydrocarbons. A transferable force field is a force field model which can be applied to almost any molecule, so long as it can be partitioned into the models available pseudo-atomic units. They are attractive since many of the interactions can be parameterised using the available data of smaller hydrocarbons and then be extended to longer hydrocarbons. Also, once a force field is parameterised, many new systems can be simulated. This allows the force field to be tested as well as being used for prediction. There are two major classes of these transferable potential models: the all-atom (AA) and united-atom (UA) models. In the AA models every atom is considered an independent interaction site, while UA models collapse the hydrogens directly connected to the carbon atoms into one pseudo-atomic unit. The use of AA models are not practical or necessary for large molecules. Thus, all the force fields used in this study are UA potential models. The UA models can be further split into two groups: carbon-centered and anisotropic. The carbon-centered models assume the interaction site of the UA is centered at the carbon of the unit whereas the anisotropic models assume that the interaction site is not on the carbon atom. As stated in the Chapter 1, the anisotropic models have not been investigated in this study. Three common transferable force fields were investigated in this study. All the parameters of the potential models used in this study may be found in Appendix A.

4.2.1 OPLS-UA

Overview

Optimized intermolecular potential functions for liquid simulations (OPLS) is one of the oldest potential models. Initially Jorgensen parameterized OPLS for united-atoms (OPLS-UA) and then later for all-atoms (OPLS-AA). Currently, OPLS has one of the largest databases of parameters, enabling it be used to predict thermodynamic properties for many compounds and mixtures. It is the *de facto* standard and its predictions

are compared in nearly every newly parameterized potential model. The focus of Jorgensen's group currently is the simulation of proteins. Jorgensen et al. (1984) introduced the OPLS-UA parameters for alkanes and alkenes. Jorgensen (1986) introduced OPLS-UA parameters for alcohols.

Implementation Details

The following pseudo-atoms were used in this study: **CH_3** (sp^3), **CH_2** (sp^3), **O**, **H**. There are in fact many other parameters available for compounds such as ethane and alkenes but since none of these have been used here their parameters will not be given.

Intramolecular interactions have been represented by:

- Constant bond lengths, Equation 4.3
- Constant bend angles, Equation 4.5
- Cosine series for the torsional potential, Equation 4.7
- LJ 12-6 potential (Equation 4.9) used for atomic-units separated by more than 3 bonds
- Coulomb potential (Equation 4.11) also used for atomic-units separated by more than 3 bonds but scaled by $\frac{1}{2}$

For the intermolecular interactions we are able write Equation 4.8 using the LJ 12-6 potential (Equation 4.9) and the Coulomb potential (Equation 4.11):

$$u^{ext} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{4.14}$$

The mixing rules used to combine the LJ parameters for OPLS-UA are very similar to the Lorenz-Berthelot (Equation 4.12) mixing rules. Instead of using an arithmetic mean for combining σ, a geometric mean is used (as is used for ϵ).

4.2.2 TraPPE-UA

Overview

Transferable potentials for phase equilibria (TraPPE) was first introduced by Martin and Siepmann (1998) as an united-atom model (TraPPE-UA). It was introduced in an attempt to overcome deficiencies in the predictions of OPLS-UA (Jorgensen et al., 1984)

and SKS (Siepmann et al., 1993). Following on from SKS, the extra parameters used for ethane were dropped. This reduced the required number of parameters and simplified its implementation. It was primarily developed to predict vapour-liquid coexistence curves (VLCC) for n-alkanes. TraPPE-UA has since been extended to predict VLCC of branched alkanes (Martin and Siepmann, 1999), branched alkenes and alkylbenzenes (Wick et al., 2000), alkanols (Chen et al., 2001). Chen and Siepmann (1999) presents an explicit-hydrogen (or all-atom) description of normal alkanes.

Implementation Details

The pseudo-atoms used are as follows: **CH$_3$** (sp^3), **CH$_2$** (sp^3), **CH$_2$** (sp^2), **CH** (sp^2), **O**, **H**. Here one can see the extra two sp^2 hybridized carbon units used for the 1-alkenes. Intramolecular interactions have been represented by:

- Constant bond lengths, Equation 4.3
- Harmonic potential for bend angles, Equation 4.6
- Cosine series for the torsional potential, Equation 4.7
- LJ 12-6 potential (Equation 4.9) used for atomic-units separated by more than 3 bonds
- Coulomb potential (Equation 4.11) also used for atomic-units separated by more than 3 bonds but scaled by $\frac{1}{2}$

Intermolecular interactions are described using the same expression as that used for OPLS-UA, Equation 4.14. The Lorenz-Berthelot (Equation 4.12) mixing rules are used to combine the LJ parameters.

4.2.3 NERD

Overview

The Nath, Escobedo and de Pablo (1998) (NERD) force field was introduced as an alternative united-atom potential model. The authors of NERD chose to have a parameter set very similar to that of the OPLS-UA. In fact, adding more parameters for propane/propene as well. As with OPLS-UA and TraPPE-UA, the NERD force field has been steadily added to. Since this study is solely interested in long-chains, these extra

parameters for short molecules are of no consequence. Branched alkanes (Nath and de Pablo, 2000; Nath and Khare, 2001), α-olefins (Nath et al., 2001*a*; Nath et al., 2001*b*), hydrogen sulfide (Nath, 2003), primary alcohols (Khare et al., 2004) have since been added. While, according to Khare et al. (2004) many other functional groups have been added to NERD, many of these have not been published or have only been presented at conferences.

Implementation Details

The pseudo-atoms used are as follows: **CH_3** (sp^3), **CH_2** (sp^3), **CH_2** (sp^2), **CH** (sp^2), **O**, **H**. Again, the extra two sp^2 hybridized carbon units have been used for the 1-alkenes. Intramolecular interactions have been represented by:

- Harmonic potential for bond lengths, Equation 4.4
- Harmonic potential for bend angles, Equation 4.6
- Cosine series for the torsional potential, Equation 4.7
- LJ 12-6 potential (Equation 4.9) used for atomic-units separated by more than 3 bonds
- Coulomb potential (Equation 4.11) also used for atomic-units separated by more than 3 bonds but scaled by $\frac{1}{2}$

The intermolecular interactions are described by the same expression as used for OPLS-UA and TraPPE-UA, Equation 4.14. The Lorenz-Berthelot (Equation 4.12) mixing rules are used to combine the LJ parameters.

Chapter 5

Simulation Techniques

Much of the basic concepts required for MC simulation have been presented. Unfortunately, there is a large gulf between MC methods and their practical implementation. This chapter focuses on the some of the most important techniques used in the practical simulation of long-chain hydrocarbons.

5.1 Random Number Generators

Today pseudorandom number generators are widely available. While, they are well established, it is important to note that they are not always very good (neither fast or random enough). The desirable properties of a particular random number generator according to James (1990) are:

1. Good distribution
2. Long period
3. Repeatability
4. Long disjoint subsequences
5. Portability
6. Efficiency

In the case of the MC simulations performed in this work, all of these properties are of great importance. Of critical importance however are the distribution and efficiency of the generators. Without a high enough level randomness a very undesirable bias can be

(sometimes unknowingly!) introduced into the MC simulations. A long period is important because of the very large number of random numbers used during the simulations, but current generators have periods of at least 10^{18} so this does not affect the simulations. Following on from this, because so many random numbers are required, the CPU time required for random number generation greatly affects the simulation times. For a detailed review and implementation of pseudorandom number generators consult James (1990), Lüscher (1994) and James (1994).

5.2 Periodic Boundaries Conditions

The time required for simulations is generally dependent on the square of the number of molecules $\mathcal{O}(N^2)$, or at best the number of molecules $\mathcal{O}(N)$. This fact can be seen simply by considering the number of intermolecular interactions, for a system of N molecules there are $\frac{1}{2}N(N-1)$ interactions. Clearly one would like to minimize the number of molecules used in a simulation. The major difficulty in doing this is the surface (or edge) effects. Even for a relatively large number of molecules, such as 1000, arranged in a $10 \times 10 \times 10$ cube, 488 molecules are still on the surface (Allen and Tildesley, 1987).

Surface effects are classically overcome by using periodic boundary conditions (PBC). The cubic box of actual molecules is replicated throughout space to form an infinite lattice. During the simulation molecules may leave the actual simulation box, enter one of the imaginary surrounding boxes, and in doing this a molecule will re-enter the simulation box from the opposite side. Figure 5.1 shows a 2-dimensional periodic system, one can see that the surrounding boxes A to H are each an exact copy of the central box. One can see that only the real simulation boxes coordinates must be stored in order to generate the whole system. The expression for the potential of this new and infinite system is:

$$u^{tot} = \frac{1}{2} \sum_{\mathbf{n}=\mathbf{0}}^{\infty} {}' \sum_{i=1}^{N} \sum_{j=1}^{N} u(|\mathbf{r}_{ij} + \mathbf{n}L|) \tag{5.1}$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, L is the periodic box size, $\mathbf{n}$ is an arbitrary 3-dimensional vector (n_x, n_y, n_z), and the prime on the $\mathbf{n}$-sum indicates that if $\mathbf{n} = \mathbf{0}$ then the $i = j$-terms are not counted. The $\mathbf{n}$ vector represents the imaginary boxes surrounding the actual simulation box. While the PBC approximation to aa infinite system works very well, it is important to note that there are some limitations:

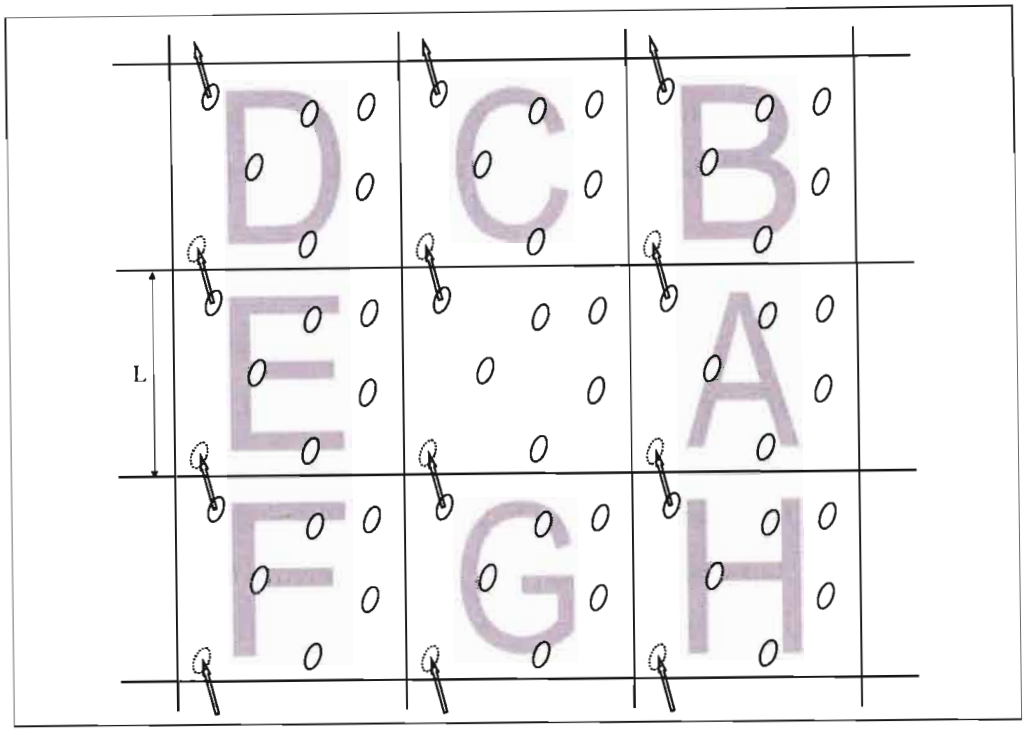

Figure 5.1: Two-dimensional periodic system (Allen and Tildesley, 1987).

- Special techniques are required when long-ranged interactions (say Coulomb interactions) are involved.
- Fluctuations that are larger than the actual simulation box cannot be represented using a periodic system (e.g. near phase transitions).
- Molecules may be able to 'sense' the symmetry of the periodic system if the boxes are made too small.

5.3 Truncation of Interactions

While the introduction of PBC solves the difficulty in dealing with the surface effects, it introduces the problem of having to determine the interactions of this much larger system. As already shown, the time required to calculate the interactions of system grows with N^2, thus additional techniques are required to handle the calculation of these interactions. The three techniques are presented in this section:

1. Minimum image convention.
2. Tail corrections.

3. Hard-inner cutoff radius.

The first two techniques reduce the number of molecules that interact with any given molecule, and the third reduces the number of unnecessary energy calculations.

5.3.1 Minimum Image Convention

The minimum image convention considers (N−1) molecules interacting with a particular particle within a box of diameter L, (i.e. within the nearest image). Note that the diameter L is the same as the simulation box length. Figure 5.2 shows the how the minimum image convention is applied to periodic system. Only the molecules within the dashed-line box (i.e. molecules from C,D,E and the central box) are considered. This effectively amounts to a truncation of the interactions within a simulated system. The circular-dash sphere is the spherical potential cutoff (r_{cut}).

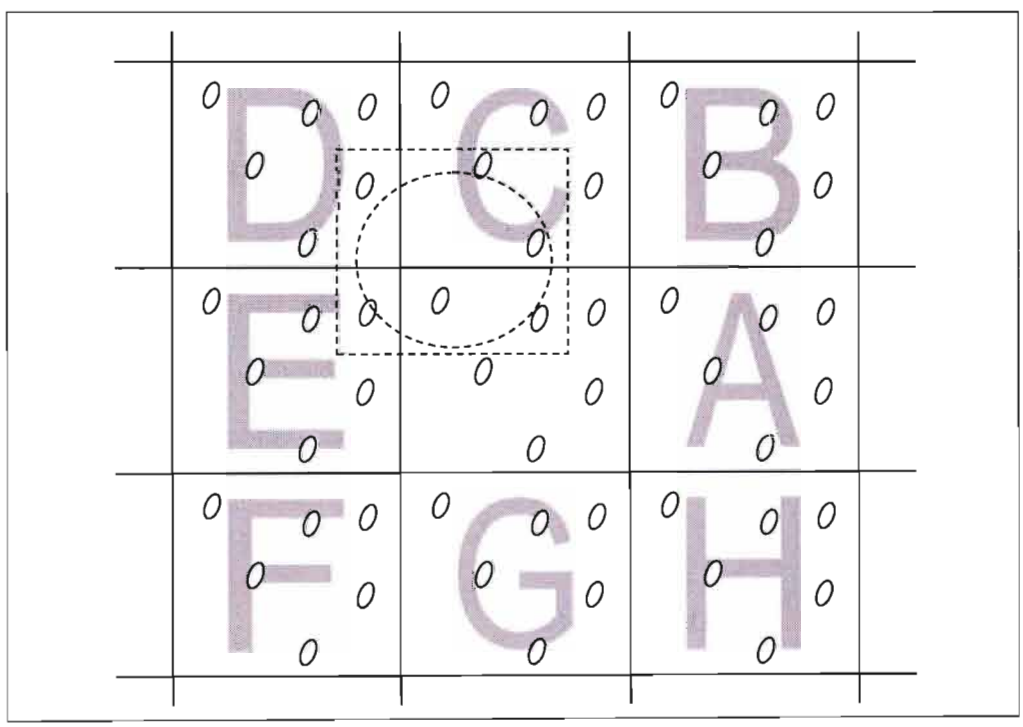

Figure 5.2: Minimum image convention and potential cutoff for a 2D periodic system (Allen and Tildesley, 1987).

5.3.2 Tail Corrections

In the case of short-range forces it is possible to split the interactions into two parts: truncated short-range interactions (typically potential energy is used in MC simulations)

and the tail correction.

$$X^{full} = X^{short} + X^{tail} \tag{5.2}$$

where X^{short} are the interactions below a cutoff radius, r_{cut} and X^{tail} the interactions above the cutoff. Below r_{cut} the interactions are calculated classically but above this radius a tail correction is used (Frenkel and Smit, 2002):

$$X^{tail} \equiv \frac{1}{2}\int_{r_{cut}}^{\infty} 4\pi\rho(r)X(r)r^2dr \tag{5.3}$$

where $X(r)$ is the interaction as a function of the radius which requires the radial distribution. It is generally assumed that above the cutoff radius the density $\rho(r)$ is equal to the bulk density ρ. Thus, Equation 5.3 may be written for a short-ranged potential as:

$$u^{tail} = \frac{N\rho}{2}\int_{r_{cut}}^{\infty} u(r)4\pi r^2dr \tag{5.4}$$

and similarly for pressure and chemical potential:

$$P^{tail} = \frac{\rho^2}{6}\int_{r_{cut}}^{\infty} r\frac{u(r)}{dr}4\pi r^2dr \tag{5.5}$$

$$\mu^{tail} = \rho\int_{r_{cut}}^{\infty} u(r)4\pi r^2dr = \frac{2u^{tail}}{N} \tag{5.6}$$

It is important to make sure that r_{cut} is large enough too ensure that the radial distribution $\mathbf{g}(\mathbf{r}) = 1$ or rather the radial density is equal to the bulk density. It is also important to make sure that r_{cut} is less than half the diameter of the periodic box. This is required to ensure that the minimum image condition is upheld (see Figure 5.2). These truncations are only of use for short-ranged forces (such as van der Waals interactions), they may not be used for long-ranged forces such as Coulomb interactions. All these tail corrections are generally calculated once off at the onset of the simulation and simply added to the appropriate property.

5.3.3 Hard-inner Cutoff Radius

Often a hard-inner cutoff radius is set for the simulations. Attempts to insert or move a molecules segment too within this cutoff distance is automatically rejected. This is done to save simulation time since these attempted insertions or moves would be rejected once the energy of the new configuration have been calculated. Typically this hard inner radius ranges from 0Å up to 2Å, the exact value of which must be chosen by checking where the radial distribution function (with no hard inner radius set) falls off to zero.

One will find that non-polar simulations will often have a larger hard-inner cutoff radius compared with the polar simulations. This makes some sense since the Coulomb attractions will override the repulsive van der Waals forces up to a point, allowing the molecules to remain closer to one-another.

5.4 The Ewald Summation

The Coulomb interactions must be considered separately to the van der Waals interactions. This is because Coulomb interactions are long-ranged, which means one cannot use the truncation of interactions method discussed previously. This fact may be shown by the direct application of Equation 5.4 to determine the tail correction for the Coulomb potential energy:

$$u_{Coul}^{tail} = \frac{N\rho}{2}\int_{r_{cut}}^{\infty} u_{coul}(r)4\pi r^2 dr = \frac{N\rho}{2}\int_{r_{cut}}^{\infty} \frac{1}{r}4\pi r^2 dr = \infty \tag{5.7}$$

Thus, one can see the necessity for some other method of handling these long-ranged forces. The potential energy due to these long-range forces must be calculable since the net charge in these systems is zero. The method most used commonly used is known as the Ewald summation. It is the technique used in both of the computer codes, TOWHEE and BIGMAC , to handle the Coulomb interactions.

A comprehensive review of the Ewald summation method is given in a three article series, viz. de Leeuw et al. (1980*a*), de Leeuw et al. (1980*b*) and de Leeuw et al. (1983). Fincham (2000) goes on to discuss the optimization of different aspects of the Ewald summation method. These articles, along with Allen and Tildesley (1987) and Frenkel and Smit (2002), discuss all the important aspects of the Ewald summation method.

5.4.1 Point Charge Ewald Summation

The following basic derivation is based on Allen and Tildesley (1987) and Frenkel and Smit (2002). Consider a system with positively and negatively charged particles (N particles) inside a cubic volume $V = L^3$. Periodic boundary conditions and an electrically neutral system ($\sum_i q_i = 0$) are assumed. The Coulomb contribution to the potential energy of this N-particle system:

$$u_{Coul} = \frac{1}{8\pi\epsilon_0}\sum_{\mathbf{n}=0}^{\infty}{}' \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|} \tag{5.8}$$

where the prime on the $\mathbf{n}$-sum indicates that when $\mathbf{n} = \mathbf{0}$ the $i = j$-terms are not counted. Equation 5.8 is conditionally convergent. What the Ewald Sum method does is effectively split Equation 5.8 into two parts: a real-space part and a fourier-space part. Basically, each point charge in the system is evenly surrounded by a charge distribution of equal magnitude but opposite charge. Gaussian distributions are used as screening distribution for each charge (Allen and Tildesley, 1987):

$$\psi(r) = q_i \kappa^3 \exp\left(\frac{-\kappa^2 r^2}{\pi^{3/2}}\right) \tag{5.9}$$

where κ is an arbitrary parameter which determines the width of the Gaussian distribution and r is the radial distance to the point charge (which is also the center of the Gaussian distribution). The screened point charges have been transformed into short-ranged interactions. These screening charge distributions must be canceled off. This has been shown graphically in Figure 5.3. It is important to note that in considering

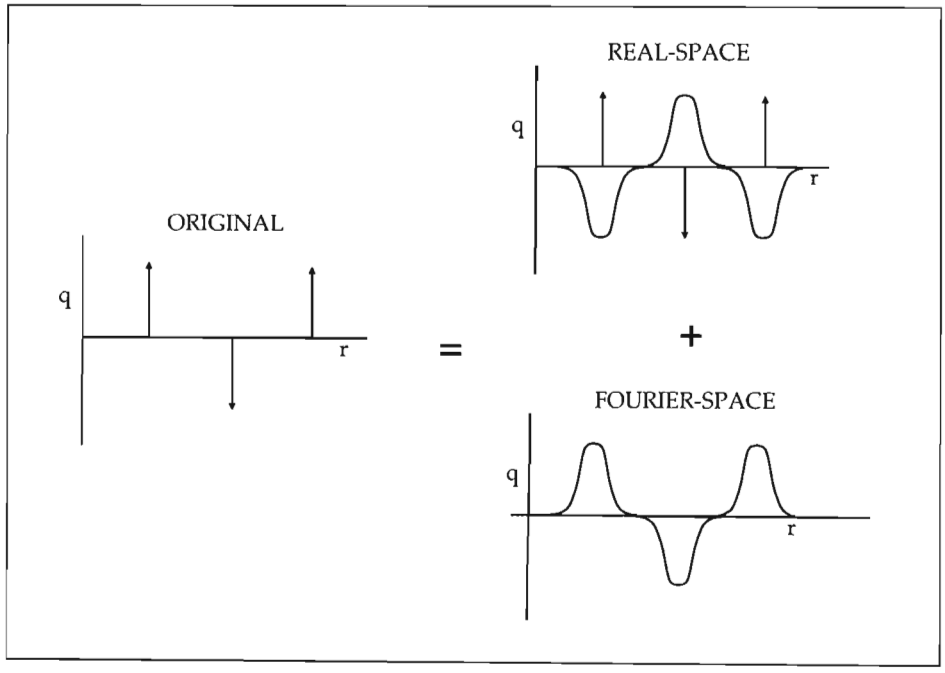

Figure 5.3: Graphical representation of screened point charges.

the charges of the system in this manner an extra, self-interaction occurs and must be corrected for. This new description of the charges in the system allows for a convergent solution of Equation 5.8. Thus, there are three terms in the Ewald Summation for point-charge systems:

1. The real-space term which includes the point charges combined with the screening distributions.

2. The fourier-space term which includes the canceling charge distributions.

3. The self-correction which accounts for the interaction of a point charge with its own screening distribution.

Detailed derivations for each of these terms may be found in Allen and Tildesley (1987) and Frenkel and Smit (2002). It can be shown that Equation 5.8 may be re-written as the following expression for the point charge Coulomb potential energy:

$$\begin{aligned} u_{Coul} = & \frac{1}{8\pi\epsilon_0}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\mathbf{n}=0}^{\infty} q_i q_j \frac{erfc(\kappa|\mathbf{r}_{ij}+\mathbf{n}L|)}{|\mathbf{r}_{ij}+\mathbf{n}L|} \\ & + \frac{1}{2V\epsilon_0}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\mathbf{k}\neq\mathbf{0}}\left(\frac{q_i q_j}{k^2}\right)\exp\left(\frac{-k^2}{4\kappa^2}\right)\cos(\mathbf{k}\cdot\mathbf{r}_{ij}) \\ & - \left(\frac{\kappa}{8\pi^{3/2}\epsilon_0}\right)\sum_{i=1}^{N} q_i^2 \end{aligned} \tag{5.10}$$

Where $erfc$ is the complementary error function, and $\mathbf{k} = 2\pi\mathbf{n}/L$. Now, from

$$\lim_{x\to\infty} erfc(x) = 0,$$

one can see that the larger κ is made the more rapidly the real-space summation will converge. Thus, if κ is set large enough then the sum over $\mathbf{n}$ will only require the sum of $\mathbf{n} = \mathbf{0}$ (i.e. the sum is reduced to the standard minimum image convention). Unfortunately, from Equation 5.9 one can see that as κ is increased so the distribution becomes sharper. This requires more terms in the fourier-space summation to account for the sharp distribution. The time required to calculate the fourier-space summation increases rapidly with the number of k-vectors used in the in the summation. Thus, one must select values for these parameters carefully in order to balance accuracy with speed. Typically the value of κ is set high enough to ensure that the real-space summation may be truncated at $\mathbf{n} = \mathbf{0}$, and the number of k-vectors is set high enough to account for the sharpness of the Gaussian distribution without requiring too much computing time. Typical values for these parameters are: $\kappa \times L = 5$ and 100-200 wave vectors (Allen and Tildesley, 1987). The number of k-vectors is generally specified by a single value k_{max}. This is the maximum number of vectors in each dimension, so the number of k-vectors in a 3-dimensional system is $(k_{max})^3$. It is very important to note that for a given simulation, $\kappa \times L$ is constant. This is done so that the sharpness of the Gaussian distribution scales with the box size. Also, the cutoff radius used for the real-space summation must

be set to half the box size. This is done to ensure that all molecules in the $\mathbf{n} = \mathbf{0}$ case are included.

5.4.2 Thermodynamic Pressure using Ewald Summation

Figure 5.4 shows the system size dependence of the virial and thermodynamic pressure of water using the SPC/E force field. One can see that the virial pressure has a much

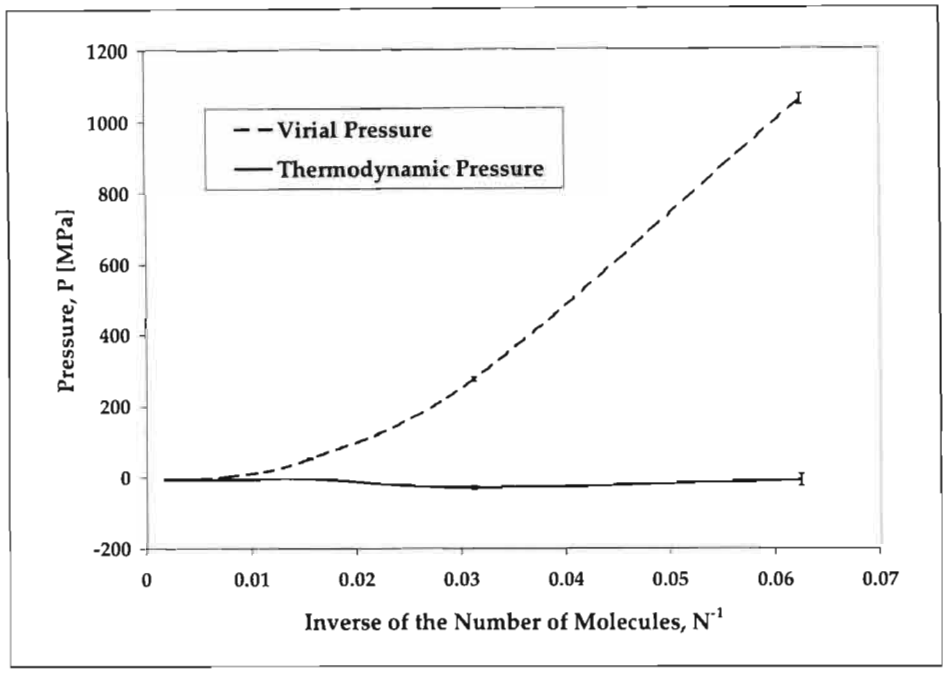

Figure 5.4: Pressure of SPC/E water as a function of the inverse number of water molecules, 1/N (Hummer and Grønbech-Jensen, 1998).

stronger size dependence compared with the thermodynamic pressure. Thus, while the two definitions of pressure are equal in the thermodynamic limit, the thermodynamic pressure is clearly superior for smaller systems.

The thermodynamic pressure is only need for the polar systems where the Ewald Sum has been used. It makes use of the following potential split:

$$U = u_{short} + u_{long-coul} \tag{5.11}$$

where u_{short} describes the ideal-gas term and the short-range pair interactions (LJ 12-6 and real-part interactions). It can be shown (Hummer and Grønbech-Jensen, 1998) that the expression for the short-range pressure is:

$$p_{short} = \rho k_B T - \frac{1}{3V} \left\langle \sum_{i=1}^{N} \sum_{j>i} r_{ij} \frac{du_{short}(r_{ij})}{dr_{ij}} \right\rangle \tag{5.12}$$

which is simply the virial pressure expression, extended to include all short-ranged potentials. It is the same for both virial and thermodynamic pressures. The interaction separation (r_{ij}) is measured between center-of-masses of two molecules, i and j. The $u_{long-coul}$ describes the long-range Coulomb interactions that are volume dependent, namely the fourier-space and self-interaction parts of the Ewald summation.

It has been shown by Hummer and Grønbech-Jensen (1998) that for polyatomic molecules the thermodynamic pressure may be expressed as follows:

$$p_{therm} = p_{short} + \frac{1}{3V}\left(\langle u_{long-coul}\rangle - \left\langle \sum_i \sum_\omega -\frac{\partial u_{long-coul}}{\partial \mathbf{r}_{i\omega}} \cdot (\mathbf{r}_{i\omega} - \mathbf{r}_i) \right\rangle\right) \quad (5.13)$$

Where $\mathbf{r}_{i\omega}$ is the position of the ω^{th} atomic site of the i^{th} molecule and $\mathbf{r}_i$ is its center-of-mass. With this definition of the pressure one is able to effectively calculate the system pressure for polar molecules when using the Ewald summation.

5.5 Configurational-bias Monte Carlo Methods

Many of methods discussed in this chapter referred to monatomic molecules. In these monatomic systems the potential energy of the system where it is fairly easy to determine. When it comes to polyatomic molecules (such as hydrocarbons) there is a need to use some method to insert these molecules and calculate the potential energy effectively. Simple, direct attempts to insert poly-atomic molecules lead to very low acceptances but it is possible to grow molecules into favorable spaces and then correct for the bias afterwards. This is the basic concept of the configurational-bias Monte Carlo (CBMC) method. Figure 5.5 shows a simplified attempt to regrow or insert the final segment of 4-unit molecule. The most favourable of the four possible trial configurations will be selected. The CBMC method, originally introduced* by Siepmann and Frenkel (1992), was partly derived from the lattice-based method presented by Rosenbluth and Rosenbluth (1955).

5.5.1 Standard CBMC

The internal potential for the hydrocarbon molecules is generally made up of the components shown in the "Intramolecular Components" (section 4.1.1). The external potential

*Later de Pablo et al. (1992*b*) introduced 'Continuum-Configurational-Bias' (CCB).

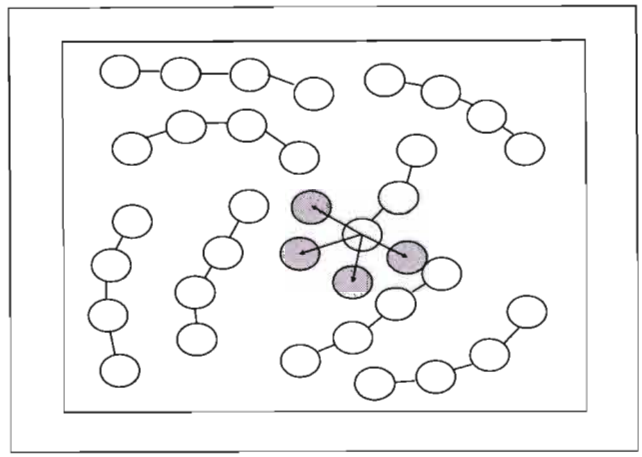

Figure 5.5: CBMC attempt to regrow last segment of 4-unit molecule.

consists of the non-bonded interactions and it it these calculations which can be very time consuming.

Now if one considers the potential energy as having two parts: an intramolecular and intermolecular part (Equation 4.1), then one may write the Rosenbluth weight of the first segment as:

$$w_1(n) = \sum_{j=1}^{j=f} \exp\left[-\beta u_{1j}^{ext}\right] \tag{5.14}$$

where f is the number of trial insertions for the first segment at random positions in the simulation box and u_{1j}^{ext} is the non-bonded potential due to the insertion of the first segment for the j^{th} attempt. A particular trial insertion for the first segment is selected using:

$$p_{1i}^{select}(\mathbf{b}_i) = \frac{\exp\left[-\beta u_{1i}^{ext}\right]}{w_1(n)} \tag{5.15}$$

For the each of the l remaining segments of the molecule, k trial orientations are generated according to the Boltzmann weight of the internal potential of that segment:

$$p_{li}^{gen}(\mathbf{b}_i)d\mathbf{b} = \frac{\exp\left[-\beta u_{li}^{int}\right] d\mathbf{b}}{\int \exp\left[-\beta u_{li}^{int}\right] d\mathbf{b}} \tag{5.16}$$

Of the k trial orientations one is selected according to the Boltzmann weight of its external potential:

$$p_{li}^{select}(\mathbf{b}_i) = \frac{\exp\left[-\beta u_{li}^{ext}\right]}{\sum_{j=1}^{j=k} \exp\left[-\beta u_{li}^{ext}\right]} \tag{5.17}$$

Each segment is selected in this manner until the entire chain has been grown. From these the Rosenbluth weight $W(n)$ of the new configuration is defined as:

$$W(n) = \frac{w_1(n) \prod_{l=2}^{l=N} \left[\sum_{j=1}^{j=k} \exp\left[-\beta u_{li}^{ext}\right]\right]}{f \times k^{N-1}} \tag{5.18}$$

where N is the number of segments in the molecule. The old configuration Rosenbluth weight $W(o)$ is determined in the same manner except that the k^{th} (or f^{th}) trial orientation is the old orientation. It is defined as follows:

$$W(o) = \frac{w_1(o)\prod_{l=2}^{l=N}\left[\sum_{j=1}^{j=k}\exp\left[-\beta u_{li}^{ext}\right]\right]}{f \times k^{N-1}} \tag{5.19}$$

Thus the original acceptance expression for particle displacements:

$$acc(o \rightarrow n) = \min\left(1, \exp\{-\beta[\mathcal{U}(n) - \mathcal{U}(o)]\}\right) \tag{5.20}$$

Becomes the following for a new configuration of a molecular (or polyatomic) system:

$$acc(o \rightarrow n) = \min\left(1, \frac{W(n)}{W(o)}\right) \tag{5.21}$$

This method can be applied to any trial move where the molecules in the system are displaced or inserted. It is unnecessary for moves affecting the volume.

5.5.2 Coupled-decoupled CBMC

The standard CBMC method is, unfortunately, limited. Vlugt et al. (1999) have shown that using the Boltzmann rejection scheme to sequentially generate bond bending angles does not result in the correct distribution of angles for branched molecules. Vlugt (2000) discusses a method in which branched segments of a molecule are grown concurrently, this is the method used in BIGMAC . The coupled-decoupled CBMC method is not absolutely required for the hydrocarbons simulated in this study, the standard CBMC is sufficient. Since it is used in the TOWHEE code and is in fact a superior CBMC method it will be briefly presented here. It is very important to note that using coupled-decoupled CBMC without the appropriate parameters will negatively impact on the simulation times. This is due to the fact that more calculations are implicitly performed per trial (with the intermolecular and torsional interactions coupled there are $n_{ext} \times n_{tor} + n_{bend} + n_{bond}$ calculations) compared with the standard CBMC method ($f + k$ calculations).

The probability of a particular configuration, generated from an appropriate trial distribution, being selected is:

$$\begin{aligned} p^{select} = \prod_{l=1}^{N} & \left[\frac{\exp[-\beta u^{ext}(i)]W_T(i)}{W_L(l)}\right]\left[\frac{\exp[-\beta u_{tors}(j)]W_B(j)}{W_T(i)}\right] \\ & \left[\frac{\exp[-\beta u_{bend}(k)]W_S(k)}{W_B(j)}\right]\left[\frac{\exp[-\beta u_{bond}(m)]}{W_S(k)}\right] \end{aligned} \tag{5.22}$$

where the corresponding Rosenbluth weights are:

$$W_L(l) = \sum_{i=1}^{n^{ext}} \exp\left[-\beta u^{ext}(i)\right] W_T(i) \tag{5.23}$$

$$W_T(i) = \sum_{j=1}^{n_{tor}} \exp\left[-\beta u_{tors}(j)\right] W_B(j) \tag{5.24}$$

$$W_B(j) = \sum_{k=1}^{n_{bend}} \exp\left[-\beta u_{bend}(k)\right] W_S(k) \tag{5.25}$$

$$W_S(k) = \sum_{m=1}^{n_{bond}} \exp\left[-\beta u_{bond}(m)\right] \tag{5.26}$$

where n^{ext}, n_{tor}, n_{bend} and n_{bond} are the number of trial sites for the intermolecular, torsional, bond bending, and bond stretching interactions respectively. The new configurations are then accepted with the following probability:

$$acc(o \rightarrow n) = min\left(1, \frac{\prod_{l=1}^{N} W_L(l)_{new}}{\prod_{l=1}^{N} W_L(l)_{old}}\right) \tag{5.27}$$

These equations have been presented with all the interactions coupled. TOWHEE does not implement this method with total coupling, rather it couples the intermolecular and torsional interactions and decouples the bond-bending and bond-stretching interactions. It is important to select the n-parameters appropriately, one must select these parameters high enough to ensure correct distribution sampling but low enough to ensure reasonable computing times. From the coupled-decoupled point of view, classical CBMC is simply the case when the intramolecular and intermolecular interactions are decoupled and all the intramolecular interactions are coupled.

5.5.3 Dual-cutoff CBMC

The dual-cutoff CBMC method (DC-CBMC), introduced by Vlugt, Martin, Smit, Siepmann and Krishna (1998), can speed CBMC simulations up by a factor of 2. DC-CBMC achieves this by reducing the time taken during the generation of a each trial orientation. It turns out that the selection process during the CBMC molecule growth phase is relatively hard-core in nature. In other words, the closest molecules affect the selection of a trial orientation substantially more than molecules further away. Thus, if one only considers the closest interactions during the molecules' growth phase and correct for this bias afterwards, then effectively the same molecular conformation will be grown in

substantially less time. Consider the external potential, u^{ext}, one may split it into two parts:

$$u^{ext} = \overline{u}^{ext} + \delta u^{ext} \tag{5.28}$$

where $\overline{u}^{ext}$ is a potential that is less expensive to calculate than u^{ext}, and δu^{ext} the difference between the two potentials. The choice of cutoff is generally made to be r_{cut^*}:

$$u^{ext} = \overline{u}^{ext}(r < r_{cut^*}) + \delta u^{ext}(r_{cut^*} < r < r_{cut}) \tag{5.29}$$

where $\overline{u}^{ext}(r < r_{cut^*})$ is a shortened potential which consists only of interactions within the distance r_{cut^*}. It is this shortened potential that is used for the generation of the molecules and because of this the Rosenbluth weights are calculated faster than when calculated classically. Using the $\overline{u}^{ext}$ to grow the molecules leads to an incorrect distribution though, thus, one must correct for this bias in the acceptance criteria. The correct acceptance rule is derived as follows:

$$\frac{acc(o \rightarrow n)}{acc(n \rightarrow o)} = \frac{\exp\left[-\beta u^{ext}(n)\right]}{\exp\left[-\beta u^{ext}(o)\right]} \frac{\exp\left[-\beta u^{ext}(o)\right]}{\overline{W}(o)} \frac{\overline{W}(n)}{\exp\left[-\beta u^{ext}(o)\right]} \tag{5.30}$$

$$= \frac{\overline{W}(n)}{\overline{W}(o)} \exp\left(-\beta\left[\delta u^{ext}(n) - \delta u^{ext}(o)\right]\right) \tag{5.31}$$

Which gives one the following, modified acceptance criteria when using DC-CBMC:

$$acc(o \rightarrow n) = \min\left(1, \frac{\overline{W}(n)}{\overline{W}(o)} \exp\left[-\beta\left[\delta u^{ext}(n) - \delta u^{ext}(o)\right]\right]\right) \tag{5.32}$$

where $\overline{W}(n)$ and $\overline{W}(o)$ are the Rosenbluth weights calculated using $\overline{u}^{ext}$.

The separation of the external potential used here is an arbitrary one. It may be split in any consistent manner as long as the interactions are independent. Figure 5.6 shows the efficiency (η, which is defined as the CPU time per number of accepted trial-moves) as a function of the second cutoff radius (r_{cut^*}) for n-octane. From Figure 5.6 one can see that the optimum value for r_{cut^*} is approximately 4.0Å. Interestingly, this value is very close to the size of the united-atoms (3.7-3.9Å). From this one can see that the controlling factor in the growth of molecules is the available space to grow into.

5.5.4 Arbitrary Trial Distributions

Martin (2004*b*) has suggested an alternative to generating trial distributions according to the Boltzmann distribution. This method has been implemented in the TOWHEE code

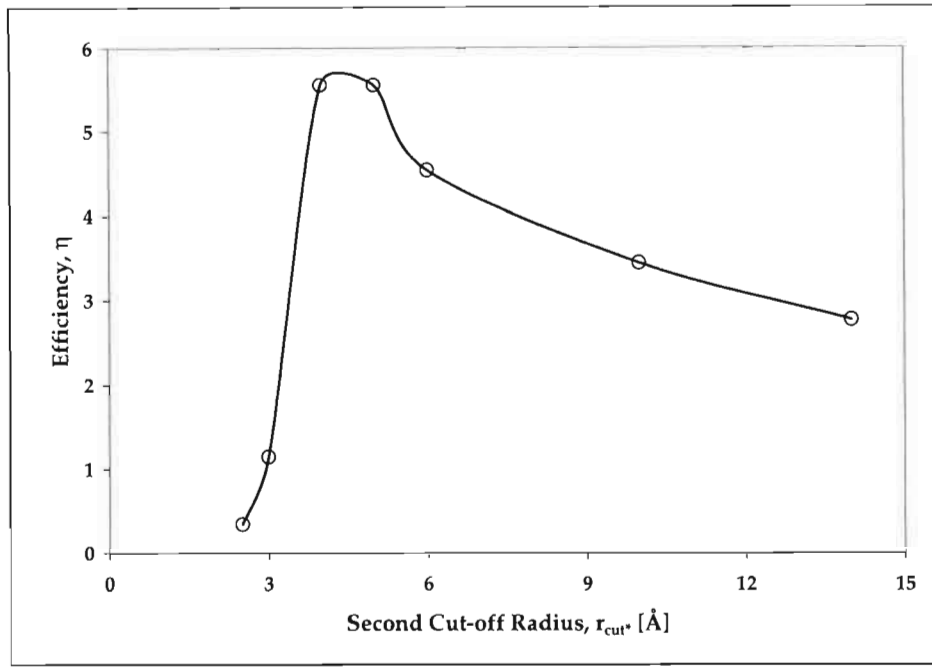

Figure 5.6: Efficiency, η (CPU time per number of accepted trial-moves), as a function of the second cutoff radius, r_{cut}, for n-octane (Vlugt, Martin, Smit, Siepmann and Krishna, 1998; Vlugt, 2000).

and increases the speed of simulations by factors of as much as two or three times. Consider the trial generation of bond lengths. Classically, bond length trials are generated according to the following probability:

$$P_{trial}^{true}(l_i) = l_0 \times \left[low^3 + rnd[0,1] \times \left(high^3 - low^3\right)\right]^{\frac{1}{3}} \tag{5.33}$$

where l_0 is the equilibrium bond length (Chapter 4), $rnd[0,1]$ is a random number between 0 and 1, and the low and $high$ are user-set values which constrain the trial lengths to reasonable lengths. Here, one must be careful not to set these limits too strictly since all bond lengths outside of low and $high$ should have effectively a zero probability of occurring. Thus, a particular trial bond length, l_i is selected based on:

$$P^{select}(l_i) = P_{trial}^{true}(l_i) \times \frac{\exp[-\beta u_{bond}(l_i)]}{W_S} \tag{5.34}$$

where W_S is the Rosenbluth weight (Equation 5.26). Now this method of generating trials is not particularly efficient since many of the generated trial lengths are unfavorable and will never be selected. The processing time involved in calculating the bond energies (and the required processing time increases for bending and torsional energy calculations) can be redirected if a more appropriate trial distribution is used. In gen-

eral:

$$P^{select}(l_i) = P^{arb}_{trial}(l_i) \times \frac{\exp[-\beta u_{bond}(l_i)] \times P^{true}_{trial}(l_i)/P^{arb}_{trial}(l_i)}{W_S} \tag{5.35}$$

where W_S (Equation 5.26) is slightly modified to:

$$W_S(k) = \sum_{m=1}^{n_{bond}} \frac{\exp[-\beta u_{bond}(m)] \times P^{true}_{trial}(l_i)}{P^{arb}_{trial}(l_i)} \tag{5.36}$$

This means that one is able to generate the trial distribution arbitrarily and then correct for it. Thus, with smart selection of the trial distribution one is able to substantially reduce the number of trial sites (n_{tor}, etc) required. To date, this has been applied exclusively to intramolecular trial configurations. It is important that the arbitrary trial distribution be nonzero over the appropriate range (positive bond lengths, bending angles between 0 and π, and dihedral angles between $-\pi$ and π) and must be straight-forward to generate from random numbers. The method used in TOWHEE is to select the trials from a Gaussian distribution, with the equilibrium values as the center means. Standard deviations must be set for each trial generation (torsional, bend and bond). As with high and low from Equation 5.33, the standard deviations must be set wide enough to not bias the trial generation. Table 5.1 shows the coupled-decoupled CBMC parameters suggested in the TOWHEE documentation. While, this method has been described and

Trial Selection	Boltzmann Distribution	Arbitrary Distribution
Bond length, n_{bond}	1000	100
Bond bend, n_{bend}	1000	100
Dihedral angle, n_{tor}	360	100
Non-bonded, n^{ext}	10	10

Table 5.1: Number of trial sites suggested in TOWHEE to be when generating trials for the coupled-decoupled CBMC algorithm.

implemented for coupled-decoupled CBMC, there is no reason why it could not be used for standard CBMC. The trial orientations for standard CBMC are generated randomly on a sphere so if only a smaller, more appropriate section of the sphere were considered then the trials generated would be far more favorable.

Chapter 6

Prediction of Phase Equilibria

Historically, the majority of Monte Carlo simulations have been involved in single-phase systems. The greatest difficulty in using these standard ensembles for multi-phases is that interfaces form and these are difficult to handle. Since phase equilibria represents such an important field of study, much research has been done in this area. Today, with the advent of methods such as the Gibbs Ensemble method (Panagiotopoulos, 1987; Panagiotopoulos et al., 1988), histogram reweighting and other pseudoensemble methods (Escobedo, 1998), the prediction of phase equilibria has come into its own. For greater detail, refer to Panagiotopoulos (2000).

The focus of this chapter is the NVT-Gibbs Ensemble method which is used for the study of single-component coexistence curves. The NVT-Gibbs method was used in this project, but some of the other popular methods, which were investigated during this study have been presented at the end of this chapter. In most cases the implementation of these alternative methods was not feasible for this study.

6.1 The NVT-Gibbs Ensemble Method

Probably the most popular method of phase equilibria determination in molecular simulation currently is the Gibbs Ensemble method (Panagiotopoulos, 1987; Panagiotopoulos et al., 1988). This technique was originally introduced as as a method to study the phase coexistence of liquid-gas systems. The introduction of this method has brought us as close to a μPT-ensemble which, as explained in Chapter 2, cannot actually be simulated due to the lack of an extensive variable to specify the system size. It involves the simulation of two linked systems (or 'boxes'), each of which represent a single phase.

There is no physical contact between these boxes, rather these boxes are linked through the three variables required for phase equilibrium (namely temperature, pressure and chemical potential). Prior to this method, each box had to either be simulated separately (using separate runs) or as a single box (which lead to many difficulties such as finite size errors or interfacial complications). The Gibbs Ensemble method has an optimum operating range between approximately $0.5T_c \rightarrow 0.9T_c$. Below about $0.5T_c$ the number of required molecule swaps attempts becomes prohibitively high due to the high liquid density, also very large vapour boxes are required due to the low vapour density. Above about $0.9T_c$ the free energy difference between the phases is insufficient to drive the system to equilibrium.

Before discussing the trial moves it is important to note that there are two subtypes of the Gibbs ensemble method. The first, NVT-Gibbs, holds the total number of molecules, the total volume, and temperature constant and the other, NPT-Gibbs, holds the pressure constant instead of the volume. Fortunately, this difference only affects the volume change move.

6.1.1 The Partition Function

As in Chapter 3, before the trial moves for an ensemble can be developed one must know the probability distribution. It is possible to construct a partition function for the Gibbs ensemble method. For a system of N particles distributed over two volumes V_1 and V_2 such that $V = V_1 + V_2$ one may write (Frenkel and Smit, 2002):

$$\begin{aligned} Q_G(N,V,T) \equiv & \sum_{n_1=0}^{N} \frac{1}{V\Lambda^{3N} n_1!(N-n_1)!} \int_0^V V_1^{n_1}(V-V_1)^{N-n_1}\,dV_1 \\ & \times \int \exp[-\beta \mathcal{U}(\mathbf{s}_1^{n_1})]d\mathbf{s}_1^{n_1} \int \exp[-\beta \mathcal{U}(\mathbf{s}_2^{N-n_1})]d\mathbf{s}_2^{N-n_1} \end{aligned} \tag{6.1}$$

And the probability distribution is proportional to:

$$\wp(n_1, V_1, \mathbf{s}_1^{n_1}, \mathbf{s}_2^{N-n_1}) \propto \frac{V_1^{n_1}(V-V_1)^{N-n_1}}{n_1!(N-n_1)!} \exp\left(-\beta[\mathcal{U}(\mathbf{s}_1^{n_1}) + \mathcal{U}(\mathbf{s}_2^{N-n_1})]\right) \tag{6.2}$$

As with the previous, classical ensembles, the acceptance expressions for the trial moves will be derived from Equation 6.2. Note that scaled coordinates have been used since both volumes change during the course of the simulation.

6.1.2 Monte Carlo Scheme

The standard Monte Carlo scheme is used here, where every cycle N moves are attempted. For the NVT-Gibbs method a trial move is selected randomly from the following possible moves:

1. Displacement of randomly selected molecule (from either box) within the same box.
2. Change the volume of each box such that the V remains constant.
3. Move a randomly selected molecule to another box.
4. Randomly selected molecule is regrown partially or totally.

It is important to remember that in the situation where one of these attempted moves is rejected, then the old configuration must be recounted (Chapter 3). The displacement move refers to positional and orientational attempted changes to a randomly selected molecule. The partial or total regrow move mentioned above is simply the application of CBMC to a molecule that is not being inserted. Depending on the simulation settings the molecule is either regrown completely, continuing on from the first unit which is left in place, or from a later unit, which is selected at random.

It is possible to assign an aspect of system equilibration to each of the moves. The molecule displacement move achieves the thermal equilibrium, the volume change move achieves mechanical equilibrium and the molecule swap move achieves the chemical equilibrium of the system.

6.1.3 Trial Moves

The trial moves will first be developed for particle systems and then extended to molecular systems using CBMC. As with the classical ensembles, the conditions of detailed balance and symmetric transition matrix α are assumed:

$$\frac{acc(o \rightarrow n)}{acc(n \rightarrow o)} = \frac{\wp(n)}{\wp(o)} \tag{6.3}$$

Thus:

$$acc(o \rightarrow n) = \min\left(1, \frac{\wp(n)}{\wp(o)}\right) \tag{6.4}$$

where n and o are the new and old configurations.

Particle Displacement

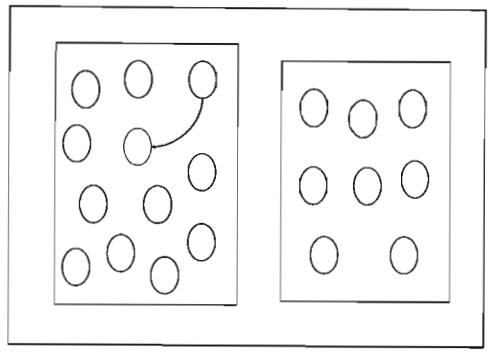

Figure 6.1: Randomly selected particle being moved within a given box.

For a particle displacement (see Figure 6.1) one can see that Equation 6.2 becomes:

$$\wp(\mathbf{s}^{n_i}) \propto \exp[-\beta \mathcal{U}(\mathbf{s}^{n_i})] \tag{6.5}$$

Where n_i refers to the number of particles in box i. Thus, simple substitution into Equation 6.4 yields the following acceptance expression:

$$acc(o \rightarrow n) = \min\left(1, \exp\left\{-\beta[\mathcal{U}(\mathbf{s}_n^{n_i}) - \mathcal{U}(\mathbf{s}_o^{n_i})]\right\}\right) \tag{6.6}$$

Which is the same as that from the standard NVT ensemble. It is appropriate to note at this point that during a particular simulation a maximum change is adjusted to maintain a particular acceptance ratio (typically ~ 0.5). This is actually in contradiction to the concept of micro-reversibility since if the maximum change is decreased then it is possible that a reverse move will not be possible. Fortuitously this does not affect the simulation substantially as long as the change in maximum displacement is not too radical.

NVT Volume Rearrangement

Figure 6.2 shows how the increase of the first box's volume decreases the second's volume. Critical to this moves is the assumption that the total volume of the system, $V = V_1 + V_2$ is conserved. In the original papers presenting the Gibbs method (Panagiotopoulos, 1987; Panagiotopoulos et al., 1988) the volume was a random walk on V_1 such that the attempted change in the volume could be written as $V_1^n = V_1^o + \Delta V$. Frenkel and Smit (2002) argue that a more appropriate random walk would be on $\ln[V_1/(V - V_1)]$ instead of on V_1. They argue that the domain of this type of walk coincides with all possible values of V_1 and and is less sensitive to the density. For this modified walk the probability

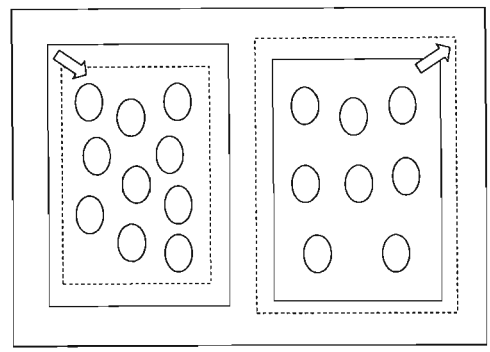

Figure 6.2: Volume change move.

distribution is proportional to:

$$\wp(\mathbf{s}^N) \propto \frac{(V_1^n)^{n_1+1}(V-V_1^n)^{N-n_1+1}}{Vn_1!(N-n_1)!}\exp[-\beta\mathcal{U}(\mathbf{s}^N)] \tag{6.7}$$

Which yields the following acceptance criteria:

$$\begin{aligned} acc(o\rightarrow n) \;=\; & \min\left\{1,\left(\frac{V_1^n}{V_1^o}\right)^{n_1+1}\left(\frac{V-V_1^n}{V-V_1^o}\right)^{N-n_1+1}\right. \\ & \left.\times\exp\left\{-\beta[\mathcal{U}(\mathbf{s}_n^N)-\mathcal{U}(\mathbf{s}_o^N)]\right\}\right\} \end{aligned} \tag{6.8}$$

As with the displacement move, a maximum volume change is adjusted to maintain a particular acceptance ratio (typically ~ 0.5). As with the displacement moves, one must be careful in how these limits are set during the simulation.

Particle Swap

Equation 6.2 may be written for the removal of a particle from box 1 and the insertion of the particle in box 2 (see Figure 6.3). The probability distribution may be written as:

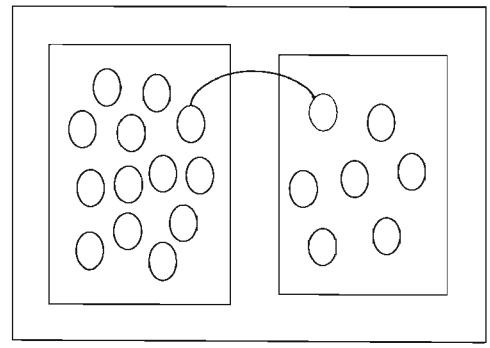

Figure 6.3: Randomly selected particle being swapped into the other box.

$$\wp(\mathbf{s}^N) \propto \frac{V_1^{n_1-1}(V-V_1)^{N-(n_1-1)}}{(n_1-1)!(N-(n_1-1))!}\exp[-\beta\mathcal{U}(\mathbf{s}^N)] \tag{6.9}$$

Which gives rise to the acceptance rule:

$$acc(o \rightarrow n) = \min\left\{1, \frac{n_1(V - V_1)}{V_1(N - n_1 + 1)} \exp\{-\beta[\mathcal{U}(\mathbf{s}_n^N) - \mathcal{U}(\mathbf{s}_o^N)]\}\right\} \tag{6.10}$$

The most important aspect of this move is to maintain a reasonable swap percentage. The higher the density (or complexity of the molecule) in a particular phase, the greater the number of attempts that will be required to achieve a particular swap percentage. Accepted exchange rates vary, but typically one modifies the fraction swap attempts to achieve approximately 1 swap per 10 cycles (Martin and Siepmann, 1998; Martin and Siepmann, 1999). In the case of this study, with the large molecules, at the lower temperatures it has proved difficult to achieve this sort exchange rate.

6.1.4 CBMC

All the moves discussed so far have been presented in terms of the potential energy of the configuration. While this is the way in which the move is developed, it is not how the move is implemented. Chapter 5 showed how CBMC can be used to greatly improve the the sampling of poly-atomic molecular systems. By applying CBMC to these moves one may obtain the following acceptance rules for the NVT-Gibbs method:

Molecule Displacement

$$acc(o \rightarrow n) = \min\left(1, \frac{W(n)}{W(o)}\right) \tag{6.11}$$

Molecule Swap

$$acc(o \rightarrow n) = \min\left(1, \frac{n_1(V - V_1)}{V_1(N - n_1 + 1)} \frac{W(n)}{W(o)}\right) \tag{6.12}$$

Note that the volume change move remains unaltered (as Equation 6.8) since the move does not change the molecule positions in the system.

6.1.5 Chemical Potential

In Appendix H of Frenkel and Smit (2002) one may find a detailed statistical mechanical investigation of the Gibbs Ensemble method. It is shown that the chemical potential of

box 1 may be defined as:

$$\mu_1 \equiv -k_B T \ln \sum_{n_1=0}^{N} \int_0^V V_1^{n_1} (V - V_1)^{N-n_1} \times \left[\frac{Q_1(n_1+1, V_1)}{Q_1(n_1, V_1)} \right] Q_2(N - n_1, V - V_1) dV_1 \quad (6.13)$$

From which the following expression for the chemical potential as an ensemble average is developed:

$$\mu_1 = -k_B T \ln \frac{1}{\Lambda^3} \left\langle \frac{V_1}{n_1 + 1} \exp\left[-\beta \Delta \mathcal{U}_1^+\right] \right\rangle_{\text{Gibbs, box 1}} \quad (6.14)$$

Where $\Delta \mathcal{U}_1^+$ is the test particle energy of a ghost particle inserted into box 1. This chemical potential is not necessary for the actual Gibbs Ensemble, rather it is used as a tool for ensuring chemical equilibrium has been reached.

6.1.6 Determination of the Critical Properties

At temperatures close to the critical temperature, $T > 0.9T_c$, the Gibbs Ensemble method begins to break down. This is due to the small difference between the two boxes Gibbs free energies, thus it becomes impossible to observe vapour-liquid coexistence. For this reason another method of determining the critical properties is required. If one does not consider the finite-size effects (this was shown to be an acceptable assumption by Panagiotopoulos (Frenkel and Smit, 2002)) then one may use the scaling law and the law of rectilinear diameters. The critical temperature is obtained by fitting the simulation results to the scaling law:

$$\rho_{liq} - \rho_{vap} = \mathcal{A}(T - T_c)^{\gamma} \quad (6.15)$$

Where ρ_{liq} and ρ_{vap} are the liquid and vapour phase densities at temperature T, T_c is the critical temperature, γ is the scaling exponent (Sometimes also know as the Ising exponent and typically represented using β), and $\mathcal{A}$ is a system dependent variable that is obtained during the fit. Once the critical temperature has been determined then it is possible to determine the critical density. Here the simulations results are fitted to the law of rectilinear diameters:

$$\frac{\rho_{liq} + \rho_{vap}}{2} = \rho_c + \mathcal{B}(T - T_c) \quad (6.16)$$

Where ρ_c, the critical density, and $\mathcal{B}$ is another system dependent variable that is obtained during the fitting procedure.

When it comes to the vapour pressure, one has many choices of the functional form to fit the data to. In this study, the Clausius-Clapeyron equation has been used:

$$\ln(P) = -\frac{\Delta H_{vap}}{RT} \tag{6.17}$$

Where P is the dimensionless pressure (P/kPa), ΔH_{vap} is the heat of vaporization, and R is the universal gas constant. The heat of vaporization is assumed to be independent of the temperature over the particular temperature range.

6.2 Other Methods

The NVT-Gibbs method played a crucial role in this study. Almost all the coexistence data produced during this study was due to this method. The following methods were investigated but did not play any major role in the coexistence data production of this study.

6.2.1 The NPT-Gibbs Ensemble Method

The NPT-Gibbs method (Panagiotopoulos et al., 1988) can be considered the Gibbs method extension for multi-component systems. The key difference is that the total volume is no longer constant, i.e. the volumes of all of the boxes may change independently. Instead of the volume, the coexistence pressure is defined. The total number of molecules ($N = N_1 + N_2$) is still conserved. All of the moves developed for the NVT-Gibbs method remain the same in the NPT-Gibbs method, expect for the volume exchange move.

NPT Volume Rearrangement

For a random walk on say V_1 such that $V_1^n = V_1^o + \Delta V$ with a pressure of P the acceptance criteria is (Panagiotopoulos et al., 1988; Panagiotopoulos, 2000):

$$acc(o \rightarrow n) = min\left(1, \left(\frac{V_i^n}{V_i^o}\right)^{n_i} \exp\{-\beta[\mathcal{U}(\mathbf{s}_n^{n_i}) - \mathcal{U}(\mathbf{s}_o^{n_i}) + P\Delta V]\}\right) \tag{6.18}$$

Where i refers the box that changes volume. While the literature presents the move for a simultaneous change in both volumes, this move is generally implemented as only changing one volume at a time.

Application Considerations

The NPT-Gibbs method may only be used for multi-component systems. This can be shown by considering the Gibbs phase rule:

$$[\text{Degrees of freedom}] = 2 - [\text{Number of phases}] + [\text{Number of components}] \quad (6.19)$$

Clearly, for a one-component, two-phase system there is only 1 degree of freedom - i.e. one may set either the temperature or the pressure but not both! Another, more practical way of thinking about this is given by Frenkel and Smit (2002). For a one-component system the two-phase region is a line in the P-T plane, thus the simulation must be specified at exactly the correct pressure and temperature in order to work! In the case of two-component systems this line becomes an area into which the temperature and pressure may be chosen (albeit, in many cases this is not much easier to achieve than in the one-component case!). Since the focus of this study was coexistence data for pure long-chain hydrocarbons, this method was only briefly investigated.

6.2.2 Histogram Reweighting

Histogram Reweighting (Ferrenberg and Swendsen, 1988; Ferrenberg and Swendsen, 1989) involves the reweighting of grand-canonical simulation results from the state (a particular μ, T) in which the simulation was performed to a new state. Figure 6.4 shows the distribution of densities for a simulation at temperature T_1 and chemical potential μ_1 reweighted to a different temperature T_2 and chemical potential μ_2. Notice that only the shape of the distribution changes, not the range of densities covered. Originally it was introduced as a method using a single (Ferrenberg and Swendsen, 1988) simulation near the critical point which was reweighted to new states near the critical point. The following year, a revised version (Ferrenberg and Swendsen, 1989) of the method was published. In this version the results of a number of simulations are combined and then reweighted to new states. This method was introduced in the late 1980's but it took about a decade for it to permeate into Monte Carlo simulation of phase equilibria. One of the earliest applications of this method, by Potoff and Panagiotopoulos (1998), showed that it may indeed be used for predicting phase equilibria. Since then work making use of Histogram Reweighting has been done on Stockmayer fluids (Kiyohara et al., 1997), Lennard-Jones mixtures (Potoff and Panagiotopoulos, 1998; Shi and Johnson, 2001; Chen

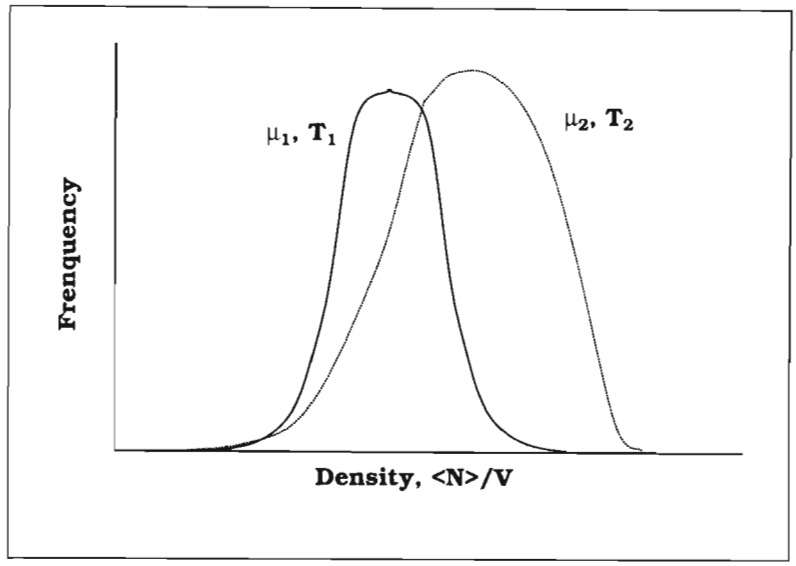

Figure 6.4: Reweighting of simulation data from T_1 to another temperature, T_2.

et al., 2001), and the Buckingham exp-6 potential (Errington and Panagiotopoulos, 1998; Errington and Panagiotopoulos, 1999*b*).

Histogram reweighting offers an effective alternative to the Gibbs method when the desired conditions are not suitable for the Gibbs method (such as low temperature or close to the critical point). It also has great potential for complex binary systems. As discussed in the NPT-Gibbs section, for non-ideal phase diagrams (where azeotropes or thin phase envelopes occur), the NPT-Gibbs simulations become very difficult, but Histogram Reweighting avoids many of these difficulties. Unfortunately, it suffers from the same major disadvantage as the Gibbs method, in that molecule insertions are required. The other disadvantage of this method is that the grand-canonical simulation results must have reasonable overlap* between one-another else the results will be very inaccurate. This is because the statistical deviations for reweighted data increases exponentially near the tails of the histograms produced. Since the method is readily extendable to multi-component systems, it will be discussed in terms of a pure-component system.

*i.e. the energies and number of molecules covered by the simulations must, in part, be the same for any two simulations

Probability Density

For a particular grand-canonical simulation, at a constant-μ, V, T, one may write the probability density as the following (Panagiotopoulos, 2000):

$$f(N,\mathcal{U}) = \frac{\Omega(N,V,\mathcal{U})\exp[-\beta(\mathcal{U}-\mu N)]}{\Xi(\mu,V,T)} \tag{6.20}$$

Where $f(N,\mathcal{U})$ is the number of occurrences of the system with N molecules and a potential energy of $\mathcal{U}$, $\Omega(N,V,\mathcal{U})$ is the microcanonical partition function and $\Xi(\mu,V,T)$ is the grand-canonical partition function. The grand-canonical partition function, Ξ, is constant for any particular grand-canonical simulation and $f(N,\mathcal{U})$ may be obtained directly from the particular simulation. Thus, if one performs a number of a grand-canonical simulations, $i = 1, 2, \ldots, R$ for at different μ_i's and T_i's but constant V, then the probability density, at a particular μ, T, may be written as:

$$\wp(N,\mathcal{U};\mu,\beta) = \frac{\sum_{i=1}^{R} f_i(N,\mathcal{U})\exp[-\beta(\mathcal{U}-\mu N)]}{\sum_{i=1}^{R} K_i \exp[-\beta_i(\mathcal{U}-\mu_i N_i) - C_i]} \tag{6.21}$$

Where K_i is the total number of observations stored during the i^{th} simulation:

$$K_i = \sum_{\mathcal{U}}\sum_{N} f_i(N,\mathcal{U}) \tag{6.22}$$

And the C_i's are the Ferrenberg-Swendsen weights which are determined from:

$$\exp(C_i) = \sum_{\mathcal{U}}\sum_{N} \wp(N,\mathcal{U};\mu_i,\beta_i) \tag{6.23}$$

Equations (6.21) and (6.23) must be iterated until the weights, C_i, have converged to within a chosen limit. This requires an initial guess for the C_i's and may take some time to converge. The probability density may be used to determine average values of certain system properties such as potential energy and density (Panagiotopoulos, 2000):

$$<\mathcal{U}>_{\mu,\beta} = \sum_{\mathcal{U}}\sum_{N} \wp(N,\mathcal{U};\mu,\beta)\mathcal{U} \tag{6.24}$$

$$<\rho>_{\mu,\beta} = \frac{1}{V}\sum_{\mathcal{U}}\sum_{N} \wp(N,\mathcal{U};\mu,\beta)N \tag{6.25}$$

Determining the Pressure

Determining the pressure at a particular μ, T is not as simple as the average density. It is possible that the virial pressure method may be used during the simulation. While this may give the pressure at the specific simulation conditions, it is of little use when

reweighting the data. The method suggested in Potoff and Panagiotopoulos (1998) makes use of the fact that at low densities the system may be considered ideal. One may then use the following expression for the pressure (Potoff and Panagiotopoulos, 1998):

$$P_{histo} = \frac{\ln \Xi(\mu, V, T)}{\beta V} + constant \tag{6.26}$$

$$= \ln \sum_{\mathcal{U}} \sum_{N} \Omega(N, V, \mathcal{U}) \exp[-\beta(\mathcal{U} - \mu N)] \tag{6.27}$$

From this one can see that the area under the probability distribution represents the partition function. For an ideal gas one may write $PV = Nk_BT$ and since it is known from Chapter 2 that $\ln \Xi = \beta PV$ then Equation 6.26 becomes $\ln \Xi + constant = N$. Now if one performs some low density simulations and extrapolates the results to $N = 0$ then the y-intercept is the additive constant from Equation 6.26.

Predicting Coexistence Curves

Two phases are in equilibrium if the temperature, pressure and chemical potentials in each phase are equal. At conditions near to phase coexistence the distribution of densities (which can be obtained by applying the reweighted probability distribution to N/V) becomes bimodal (Figure 6.5). Since the areas under each of these curves gives the pressure in that phase. Thus the basic procedure one could follow in order to determine the coexistence curve is:

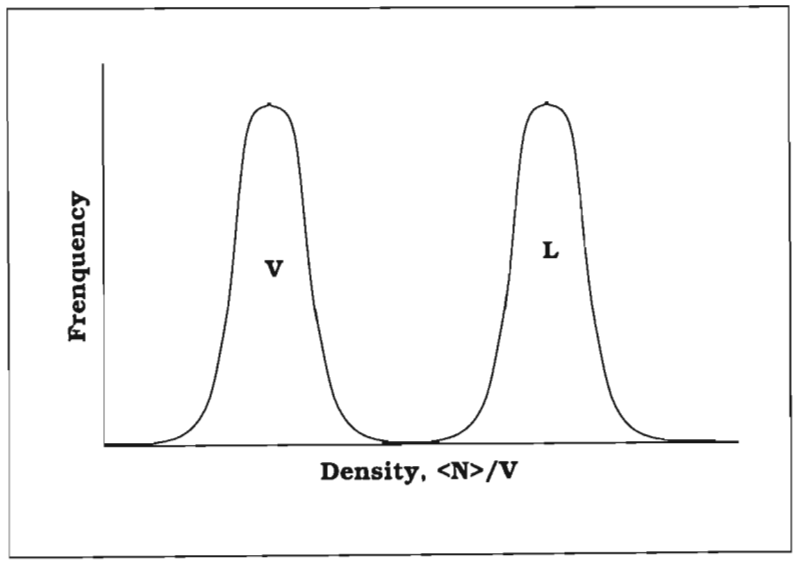

Figure 6.5: Bimodal distribution of densities at a given μ, V, T indicating phase separation.

- Select a range of temperatures to determine coexistence at;

- Reweight the simulation data to these temperatures;
- Find the chemical potential at which phase coexistence occurs (using say Newton-Raphson)

It is important to explore enough of the phase space during the grand-canonical simulations since if there is insufficient overlapping of the energy and density distributions the results will be questionable.

Extending to Multi-component Systems

Histogram reweighting may be extended to any number of components, for example a binary system:

$$f(N_1, N_2, \mathcal{U}) = \frac{\Omega(N_1, N_2, V, \mathcal{U}) \exp[-\beta(\mathcal{U} - \mu_1 N_1 - \mu_2 N_2)]}{\Xi(\mu_1, \mu_2, V, T)} \tag{6.28}$$

The only limit to this is the complexity (and practicality) in dealing the multi-dimensional histograms that such systems produce. One can see that the dimensionality increases at components + 1, thus a ternary system would produce a 4-D histogram. Generally these higher order histograms are stored as lists, but as computational power of modern computers increase it may well prove viable to use multi-dimensional arrays.

6.2.3 Gibbs-Duhem Integration

The Gibbs-Duhem Integration method (Kofke, 1993*b*; Kofke, 1993*a*) was investigated as a possible method due to the fact that molecule insertions were not required. This fact alone made the method attractive, unfortunately there were a number of other issues which made its use impractical. The greatest hinderance being the fact that a point on the coexistence curve was required *a priori*. And if a point was determined, using say the Gibbs method, then any inaccuracies in this point would propagate throughout the rest of the predicted curve. For pure-component systems the Clausius-Clapeyron equation is used in the following form (Kofke, 1993*b*):

$$\left(\frac{dP}{d\beta}\right)_{sat} = -\frac{\Delta H}{\beta \Delta V} \tag{6.29}$$

Or, more appropriately, if one takes the natural logarithm of the pressure:

$$\left(\frac{d \ln P}{d\beta}\right)_{sat} = -\frac{\Delta H}{\beta P \Delta V} = f(\beta, P) \tag{6.30}$$

Where ΔH is the difference in enthalpies of the two coexisting phases, and ΔV is the difference in molar volumes between the two phases. Equation 6.30 is the form generally used and the ΔH and ΔV are obtained from NPT MC simulations. Predictor-corrector methods of numerical integration are used to minimize the number of function evaluations. An added improvement was the simultaneous changing of the volumes of the two independent NPT simulations. This meant that $f(\beta, P)$ could be re-evaluated continually based on the running averages for the volumes and enthalpies of the two phases and the result used to update the estimate of the pressure, all during the simulation (Kofke, 1993*b*).

As explained, the requirement of a starting point, the uncertainty in the subsequent results and fact that this method also suffers from instability near the critical point lead us to forgo is use in this study. For further information on this method, one may refer to the original papers (Kofke, 1993*b*; Kofke, 1993*a*) or to reviews by a number of different authors (Escobedo, 1998; Escobedo, 1999; Panagiotopoulos, 2000).

Chapter 7

Simulation Results

Three groups of linear hydrocarbons were investigated, viz. n-alkanes, 1-alkenes, and 1-alcohols. Three force fields, OPLS-UA, NERD and TraPPE-UA were used to simulate these long-chain hydrocarbon compounds. The properties obtained from the simulations have been presented in this chapter are the coexistence densities and vapour pressures. All the literature data used in this chapter may be found in Appendix D.

7.1 Simulation Details

The simulations undertaken during this study can be split into two groups: non-polar and polar simulations. This split is appropriate since two different codes were used for each type. The BIGMAC code was used for the non-polar simulations; while the TOWHEE code was used for the polar simulations. The reason for this was simply that BIGMAC had only been developed for non-polar molecules while TOWHEE can simulate both polar and non-polar compounds. Both codes are similar in many respects but there are still a number of important differences. The major difference is that TOWHEE makes use of coupled-decoupled CBMC whereas BIGMAC uses standard CBMC. Further information about these two simulation codes may be found in Appendix B. The non-polar simulations encompass the n-alkanes, 1-alkenes and binary n-alkane mixtures. The 1-alcohols were the only polar compounds simulated during this study.

7.1.1 Non-polar Simulations

Pure-Component Systems

NVT-Gibbs ensemble Monte Carlo simulations were performed in order to calculate the vapour-liquid coexistence curves for the pure-component non-polar compounds. The number of molecules used in each simulation varied with the length of the hydrocarbon compound. They ranged from 300 molecules for the C_8 simulations down to about 100 molecules for the C_{20} simulations. Selection of systems sizes was based on already published simulation results and finite-size testing (discussed at length in Chapter 8). Two force fields were used for the non-polar simulations, NERD and TraPPE-UA, the implementation details of which may be found in Chapter 4, and model parameters may be found in Appendix A. These potential models use the Lennard-Jones (LJ) 12-6 potential to describe the non-bonded intramolecular and intermolecular interactions. In all simulations the LJ potential was truncated after 14Å and the standard analytic tail corrections were enforced. An additional center-of-mass based cutoff was used where all attempted insertions or displacements within 2Å of another molecule were automatically rejected. Standard configurational bias Monte Carlo (CBMC) was used in these simulations. The number of trial insertions used for the CBMC method varied with hydrocarbon length from 10 trials for C_8 up to approximately 20 trials for C_{20}. Dual-cutoff CBMC was implemented with a cutoff radius of 5Å. The following fixed probabilities were generally used: for volume exchanges (0.006), molecule swaps (0.328), translational (0.222), rotational (0.222) and conformational* (0.222) moves. If the number of accepted insertions dropped too low then the molecule swaps fraction was increased at the expense of the displacement moves. The simulations were first run for at least 30000 cycles to allow the systems to reach equilibrium. Thereafter, the simulations were run for another 80000 to 120000 cycles. It was these production runs that were used to determine the average coexistence densities and vapour pressures for a specified temperature. The standard deviations were computed by splitting the production runs into 5 blocks. The temperatures at which the simulations were performed was based on the normal boiling points and critical temperatures. Basically four to six temperatures were selected over this range, from slightly below the normal boiling point to about $0.9T_c$.

*These are partial and total CBMC molecule regrowths

Binary Systems

Simulation of multi-component systems requires two stages. The pure-component properties are determined as described above, using NVT-Gibbs method. The NPT-Gibbs method is then used for the binary mixture simulations. The NPT-Gibbs simulations yield the average liquid-vapour compositions and densities.

Most of the simulation settings used for the NPT-Gibbs method simulations for this study are the same as for the pure-component simulations described above. The most important aspect of the NPT-Gibbs method is that instead of the volume being specified, the system pressure is specified. The difficulties inherent in using this method have been discussed in Chapter 6. Raoult's law was used to estimate the system pressure so that the simulation conditions would be initially specified within the phase envelope.

7.1.2 Polar Simulations

As with the pure-component non-polar simulations, the NVT-Gibbs method was used to generate the vapour-liquid coexistence curves. Here the systems sizes ranged from about 200 molecules for octanol (C_8) down to about 100 molecules for eicosanol (C_{20}). Again, these system sizes were primarily based on published literature. Three potential models were used to simulate the primary alcohols: NERD, TraPPE-UA and OPLS-UA. The OPLS-UA force field also uses the LJ 12-6 potential to describe the non-bonded intramolecular and the van der Waals intermolecular interactions. As with the non-polar simulations a cutoff of 14Å was used with standard analytical tail corrections. The additional center-of-mass based cutoff was used for some of the shorter alcohol simulations but not for the longer alcohols. The size of the cutoff, when used, was about 1Å. The Coulomb potential was used to describe the electrostatic intermolecular interactions. This long-ranged potential was handled using Ewald sums with tin-foil boundary conditions ($\epsilon_s = \infty$). The parameters used for the Ewald sum were: $\kappa \times L = 5$ and $k_{max} = 5$. Coupled-decoupled CBMC was used for these simulations. With the implementation of the coupled-decoupled CBMC method, many more parameters had to be set (and optimized). Arbitrary trial distributions were used to help speed up the CBMC for all intramolecular terms. Table 7.1 shows the normal parameters used during the alcohol simulations of this study. For the longer alcohols these values were increased (up to 20 for C_{20}) to maintain the the number of swap moves accepted. Dual-cutoff CBMC was

Trial Selection	Parameter Value
Bond length, n_{bond}	10
Bond bend, n_{bend}	10
Dihedral angle, n_{tor}	10
Non-bonded, n^{ext}	10

Table 7.1: Standard number of trial sites used for coupled-decoupled CBMC in the primary alcohol simulations.

again implemented using a cutoff radius of 5Å. The standard fixed probabilities for trial moves were: for volume exchanges (0.006), molecule swaps (0.328), translational (0.222), rotational (0.222) and conformational (0.222) moves. As with non-polar simulations, the molecule swaps were adjusted to maintain sufficient molecule swaps. Equilibration runs were for at least 30000 cycles and the production runs were for another 50000 to 80000 cycles. Standard deviations for these simulations were calculated by splitting the production runs into 5 to 10 blocks. The temperatures at which the simulations were performed are determined in the same way as for the non-polar simulations.

7.2 Simulation Results

7.2.1 n-Alkanes

n-Alkanes were the first group of hydrocarbons investigated in this study. The n-alkanes investigated were n-octane, n-hexadecane, n-tetracosane. Only NERD and TraPPE-UA force fields were used since their superiority over OPLS-UA has already been established in published literature (Martin and Siepmann, 1998; Nath, Escobedo and de Pablo, 1998). The density coexistence curves are shown in Figure 7.1. The vapour pressures are shown in Figure 7.2 using a Clausius-Clapeyron (CC) plot. For clarity the standard deviations of the simulation results have not been included in the graphs. Typically there was a 1–3% deviation in liquid densities and a 10–15% deviation in the vapour densities and pressures. Numerical results (with the standard deviations) may be found in Appendix C. The results follow the available experimental and predicted curves very well for n-octane and n-hexadecane. It appears that the predicted coexistence curve for n-tetracosane may be slightly shallow. Both NERD and TraPPE-UA for all the n-alkanes consistently over-predicted the vapour pressures slightly.

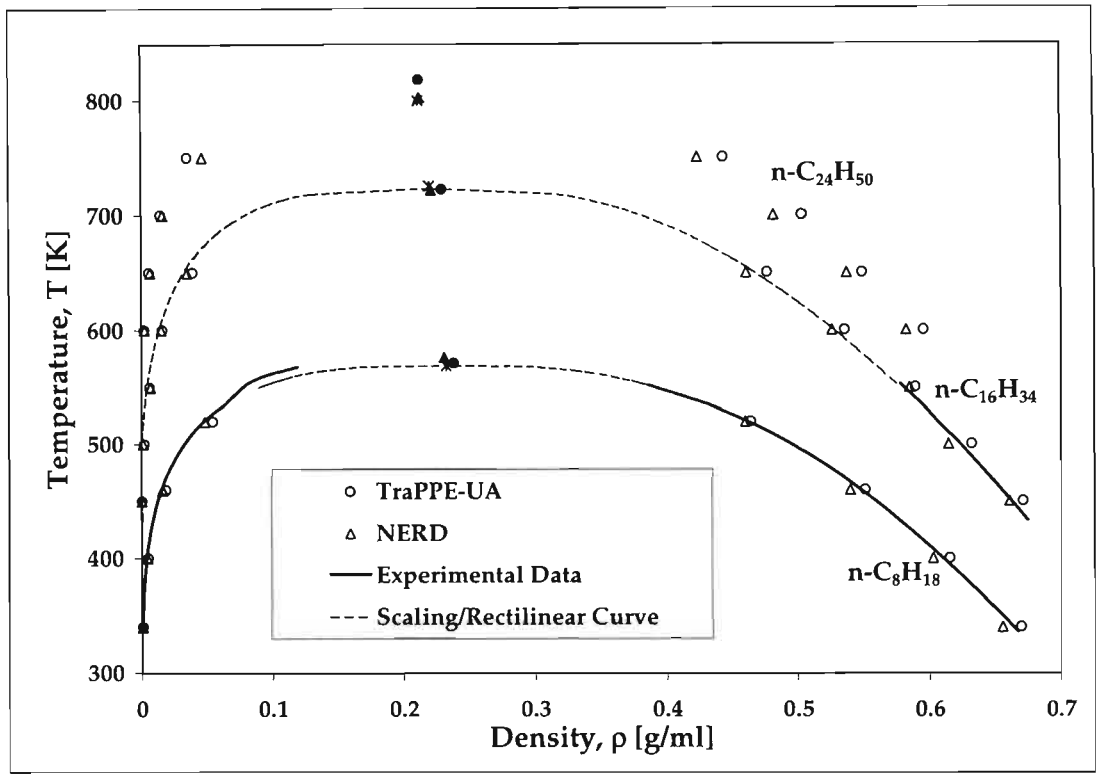

Figure 7.1: Coexistence density plots for the n-alkanes. Filled triangles (NERD), circles (TraPPE-UA) and stars (Experimental) are the critical properties.

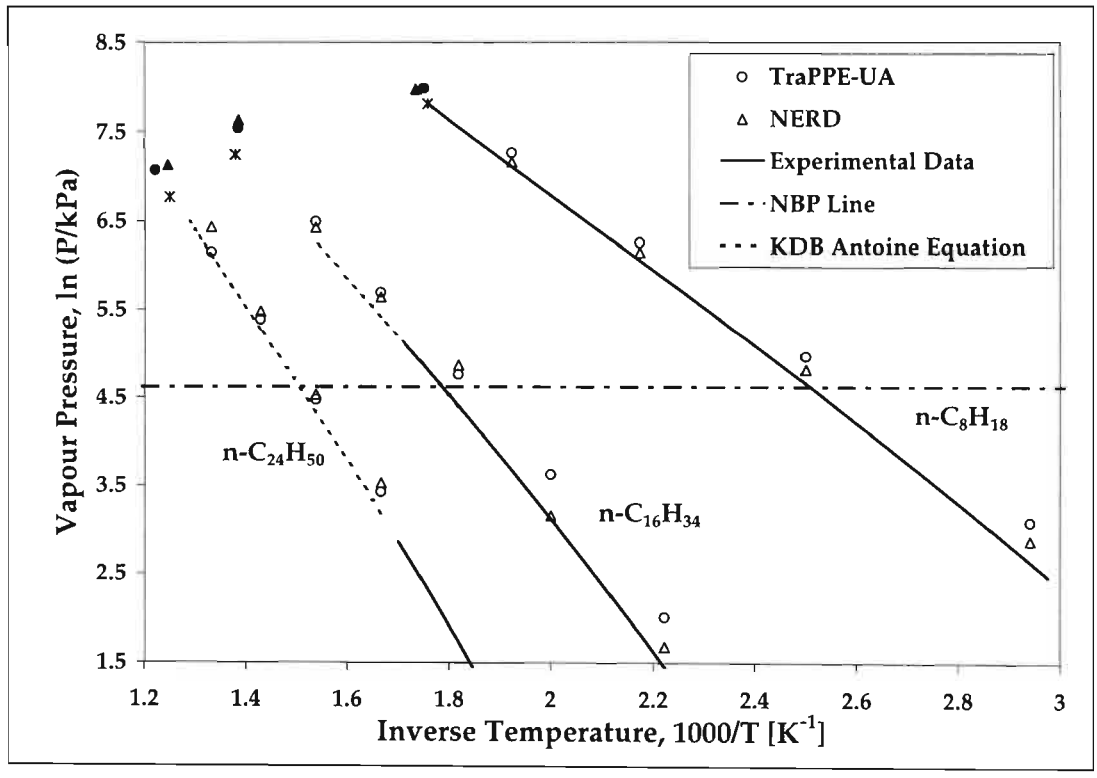

Figure 7.2: Vapour pressure plots for the n-alkanes. Filled triangles (NERD), circles (TraPPE-UA) and stars (Experimental) are the critical properties.

7.2.2 1-Alkenes

1-Alkenes were the second group of hydrocarbons investigated in this study. The 1-alkenes investigated were 1-octene, 1-decene, 1-hexadecene. As with the n-alkanes, only NERD and TraPPE-UA force fields were used. Figure 7.3 shows the coexistence densities for the simulated 1-alkenes and Figure 7.4 shows the vapour pressures using a CC plot. Numerical results may also be found in Appendix C. As with the n-alkanes, the standard deviations have not been shown but were typically 1–3% for the liquid densities and 10–15% for the vapour densities and pressures. While there is a lack of experimental density data for the 1-alkenes, the results for the 1-decene and 1-hexadecene follow similar trends as the results for shorter 1-alkenes such as 1-octene. As for the n-alkanes, the vapour pressures are in good agreement with available experimental and predicted data, but they are still consistently over-predicted slightly.

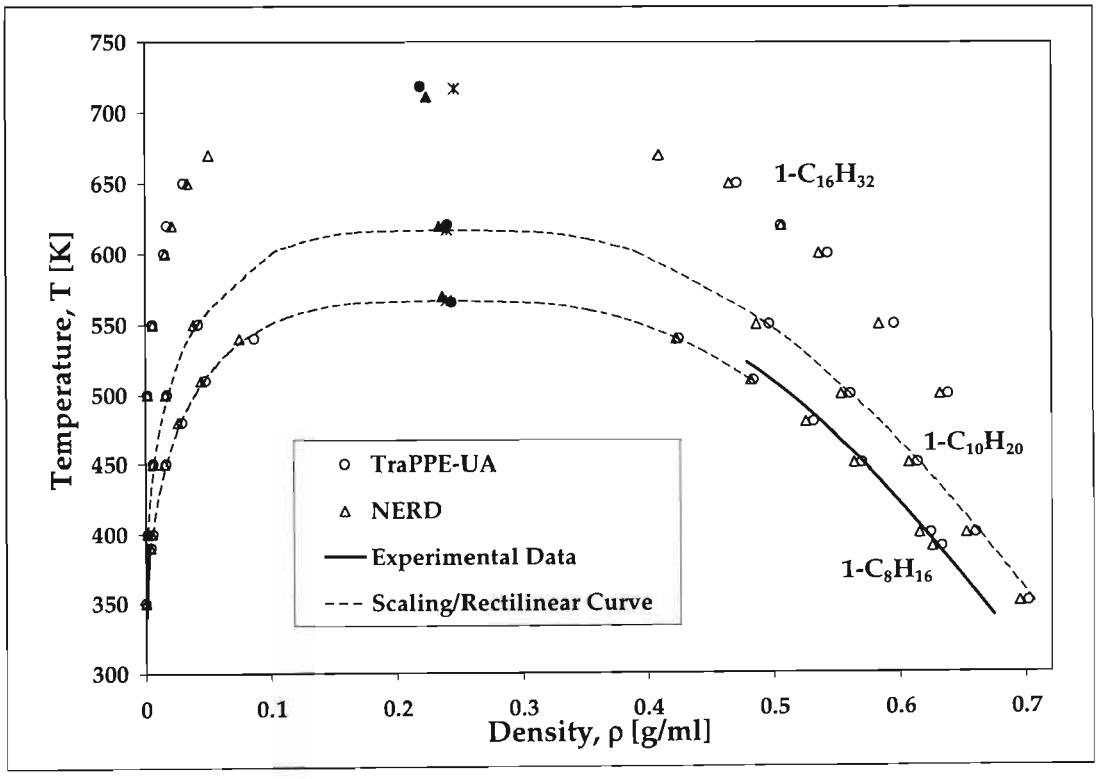

Figure 7.3: Coexistence density plots for the 1-alkenes. Filled triangles (NERD), circles (TraPPE-UA) and stars (Experimental) are the critical properties.

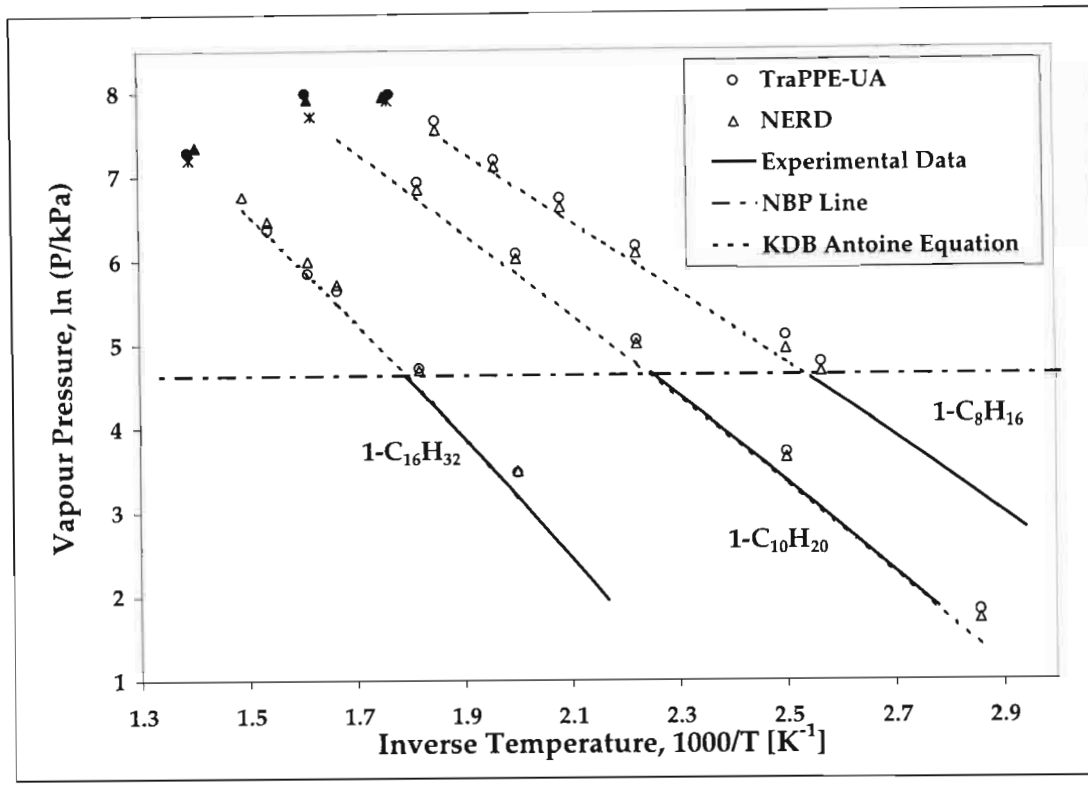

Figure 7.4: Vapour pressure plots for the 1-alkenes. Filled triangles (NERD), circles (TraPPE-UA) and stars (Experimental) are the critical properties.

7.2.3 1-Alcohols

The final group of hydrocarbons investigated in this study were the primary alcohols. The 1-alcohols investigated were 1-octanol, 1-dodecanol, 1-hexadecanol and 1-eiosanol. All three force fields, OPLS-UA, NERD and TraPPE-UA were used to simulate these alcohols. Due to the large difference between the predicted densities and vapour pressures for the alcohols, different graphs have been used for each alcohol investigated. Figures 7.5 (C_8), 7.7 (C_{12}), 7.9 (C_{16}) and 7.11 (C_{20}) show the coexistence densities. Figures 7.6 (C_8), 7.8 (C_{12}), 7.10 (C_{16}) and 7.12 (C_{20}) show the vapour pressures using CC plots. The numerical results may be found in Appendix C. As with the n-alkanes and 1-alkenes, the standard deviations have not been shown but were typically 1–3% for the liquid densities and 10–15% for the vapour densities and pressures. In all alcohol simulations OPLS-UA over predicts the liquid densities and under predicts the vapour densities and pressures substantially. Both NERD and TraPPE-UA perform better than OPLS-UA, but still generate shallow coexistence curves and consistently over predict the vapour pressures. The plots for 1-eicosanol are likely to follow the same trends.

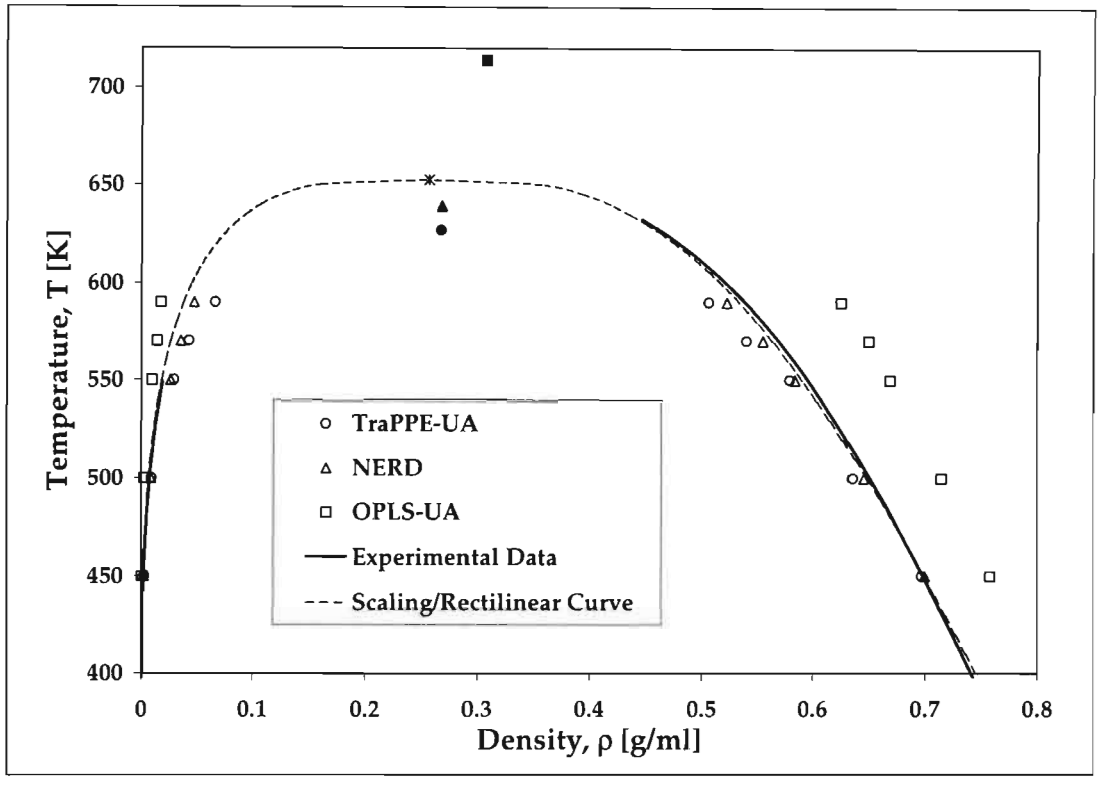

Figure 7.5: Coexistence density plots for 1-octanol. Filled triangles (NERD), circles (TraPPE-UA), squares (OPLS-UA) and stars (Experimental) are the critical properties.

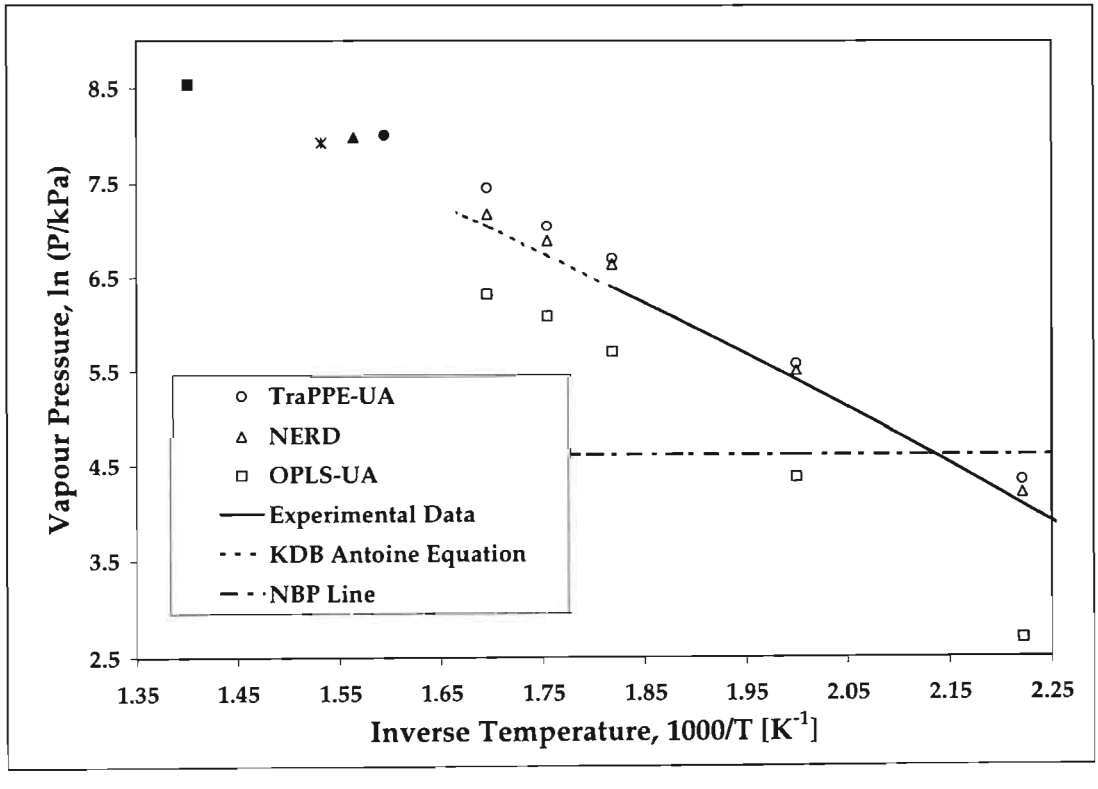

Figure 7.6: Vapour pressure plots for 1-octanol. Filled triangles (NERD), circles (TraPPE-UA), squares (OPLS-UA) and stars (Experimental) are the critical properties.

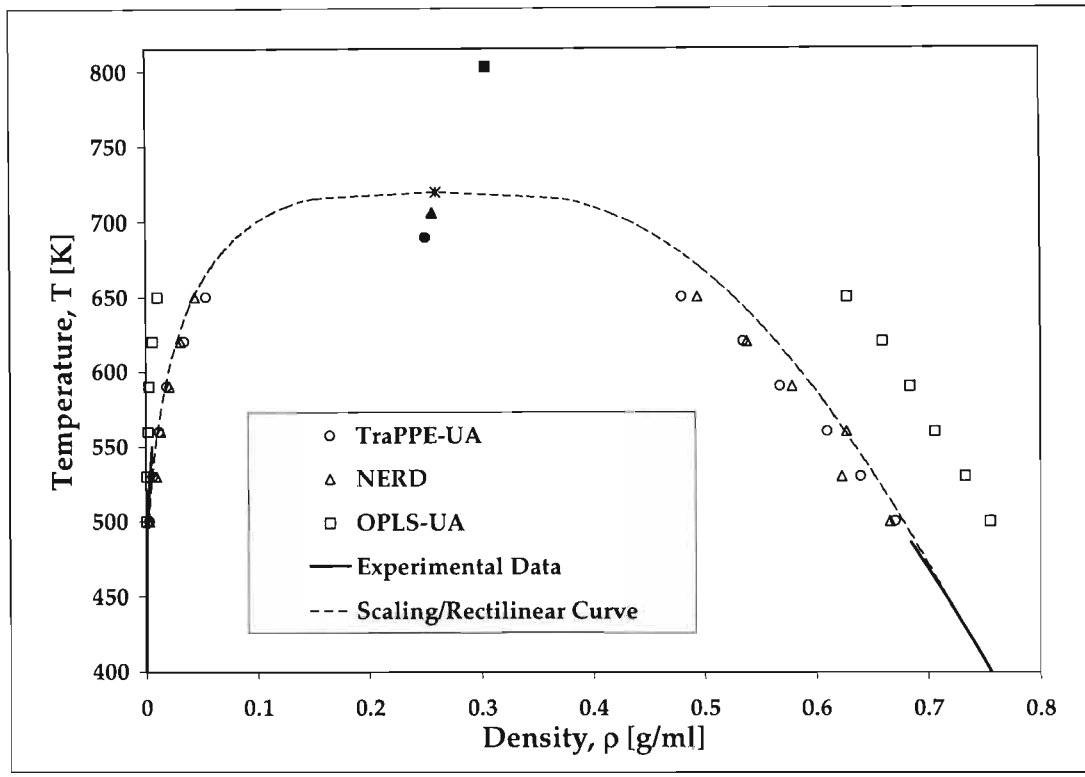

Figure 7.7: Coexistence density plots for 1-dodecanol. Filled triangles (NERD), circles (TraPPE-UA), squares (OPLS-UA) and stars (Experimental) are the critical properties.

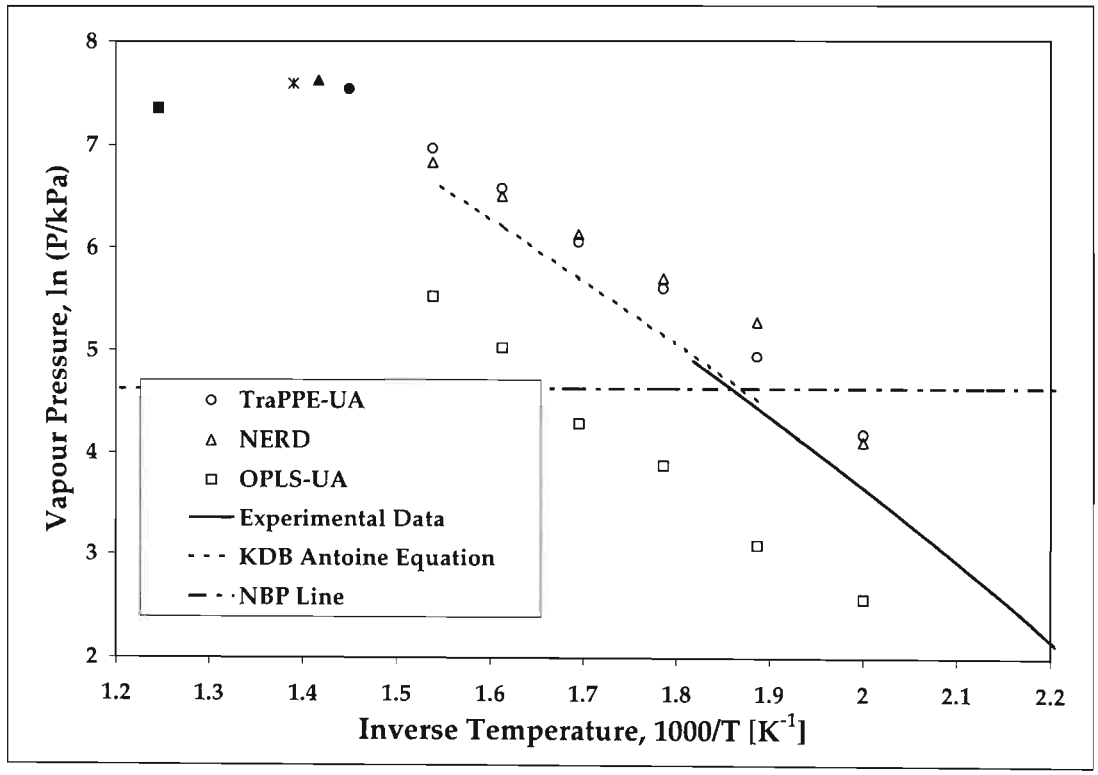

Figure 7.8: Vapour pressure plots for 1-dodecanol. Filled triangles (NERD), circles (TraPPE-UA), squares (OPLS-UA) and stars (Experimental) are the critical properties.

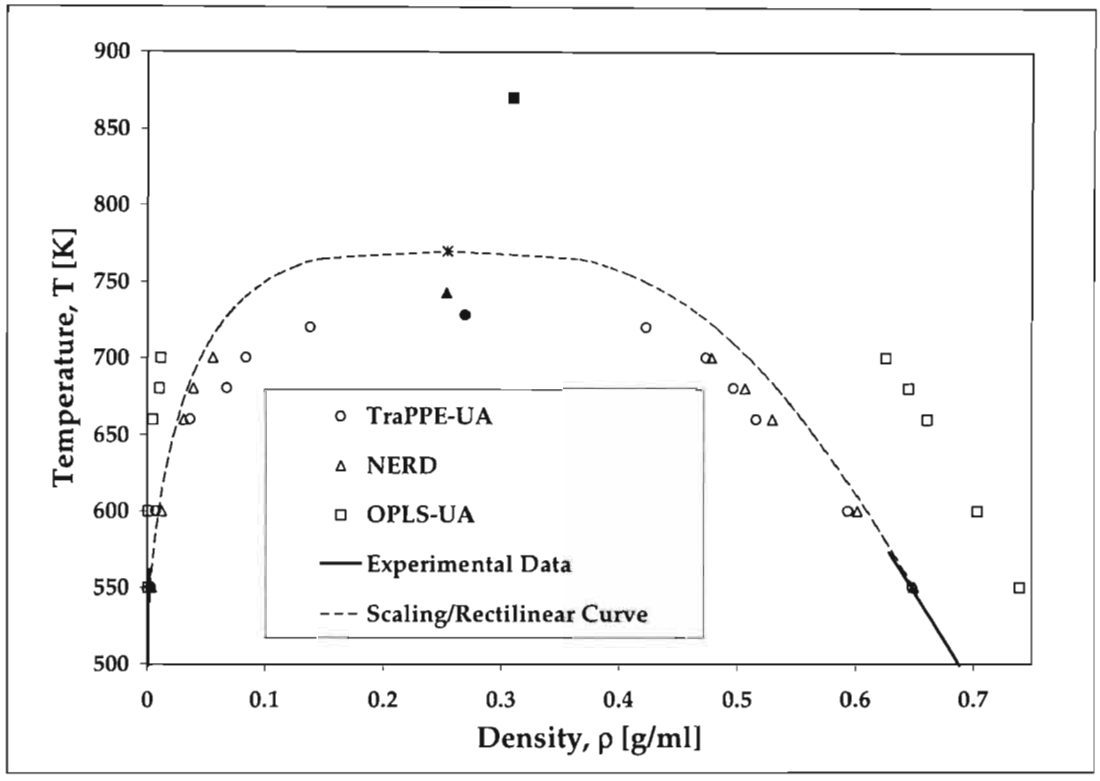

Figure 7.9: Coexistence density plots for 1-hexadecanol. Filled triangles (NERD), circles (TraPPE-UA), squares (OPLS-UA) and stars (Experimental) are the critical properties.

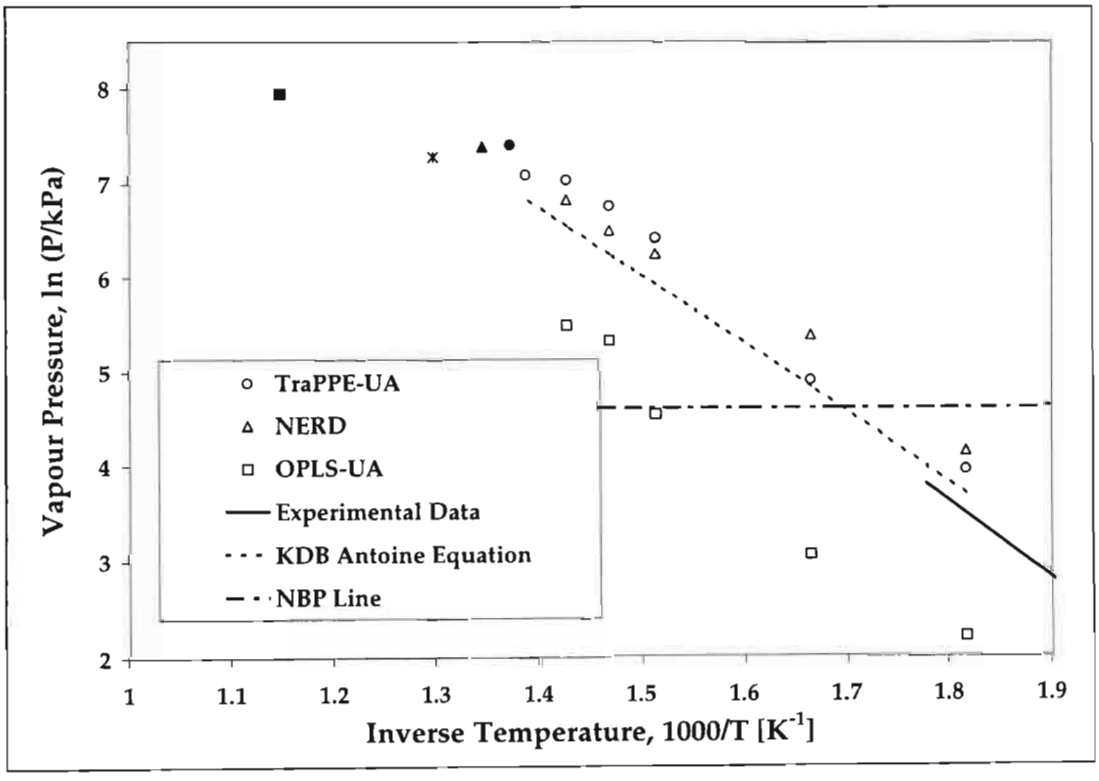

Figure 7.10: Vapour pressure plots for 1-hexadecanol. Filled triangles (NERD), circles (TraPPE-UA), squares (OPLS-UA) and stars (Experimental) are the critical properties.

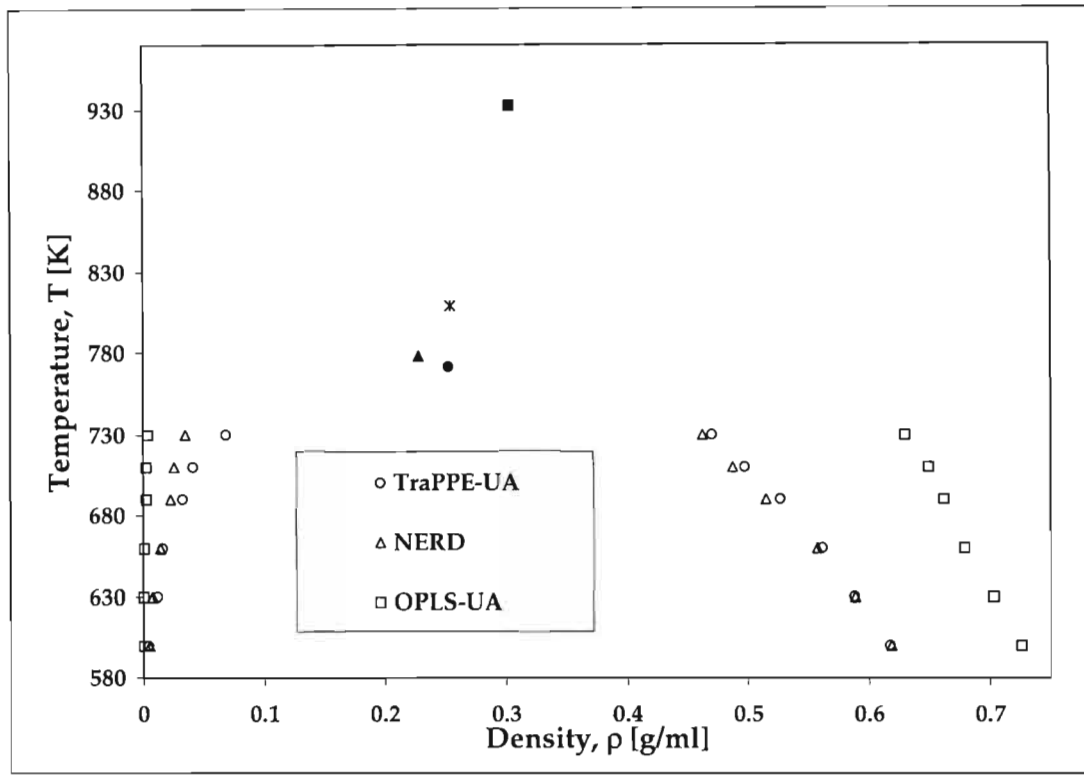

Figure 7.11: Coexistence density plots for 1-eicosanol. Filled triangles (NERD), circles (TraPPE-UA), squares (OPLS-UA) and stars (Experimental) are the critical properties.

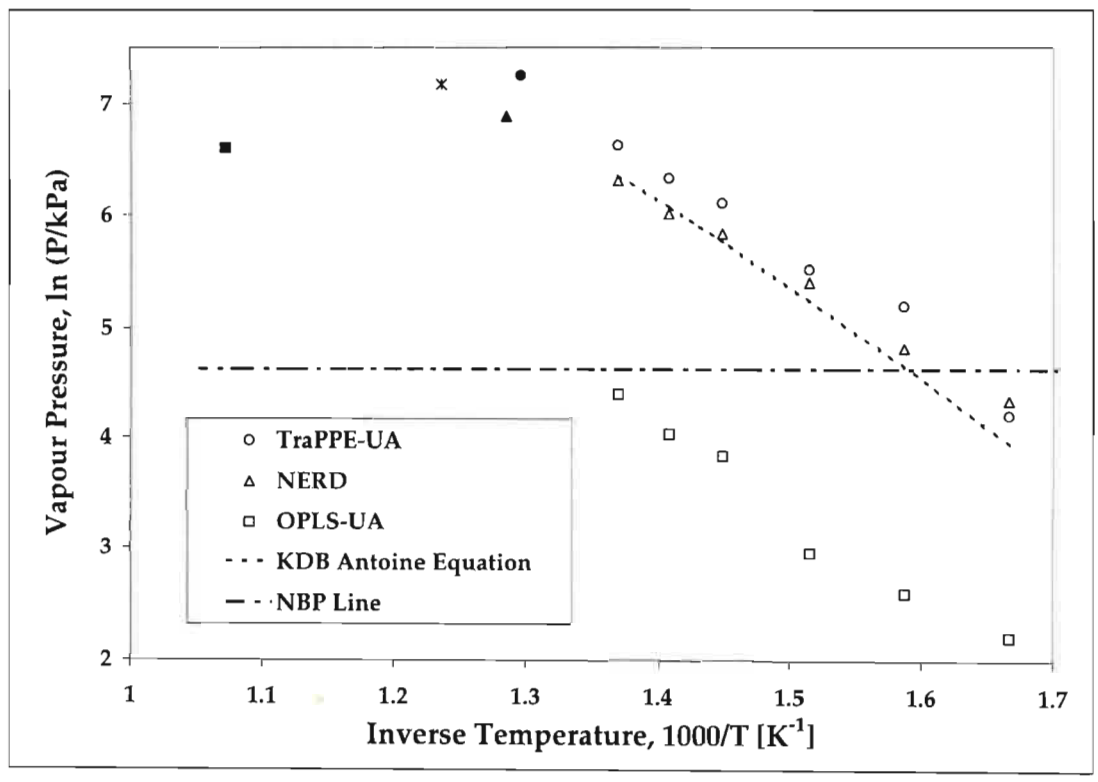

Figure 7.12: Vapour pressure plots for 1-eicosanol. Filled triangles (NERD), circles (TraPPE-UA), squares (OPLS-UA) and stars (Experimental) are the critical properties.

7.2.4 Binary n-Alkane Mixtures

The following two n-alkane mixtures were briefly investigated using the NERD force field:

- n-pentane(1) + n-octane(2), at 308.7K
- n-hexane(1) + n-hexadecane(2), at 289.15K

For the n-pentane + n-octane system the x-y plot and P-x-y plot are shown. For the n-hexane + n-hexadecane system only the P-x-y plot has been shown since there is no experimental data for the vapour composition to be compared to. As with the previous results the typical standard deviation was 10% for the vapour compositions and pressures and 1% for the liquid compositions. Both systems were predicted with good accuracy, although the P-x curve for the n-hexane + n-hexadecane system was slightly high.

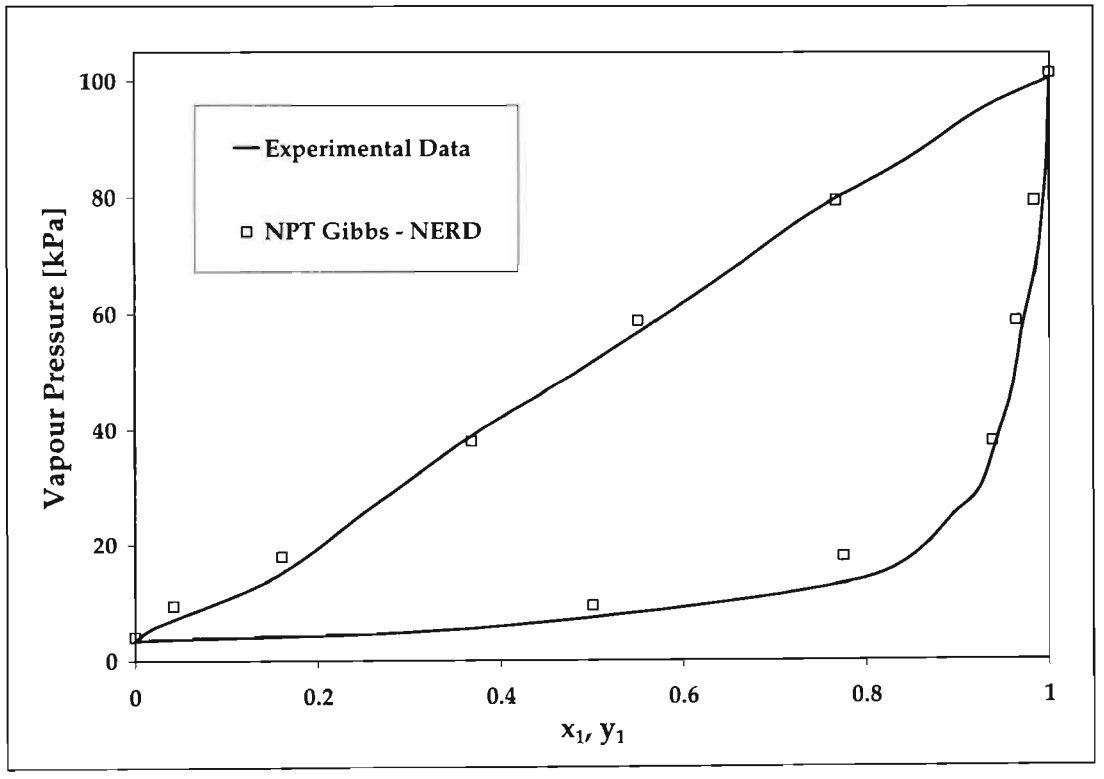

Figure 7.13: P-x-y plot for n-pentane(1) + n-octane(2) mixture at 308.7K.

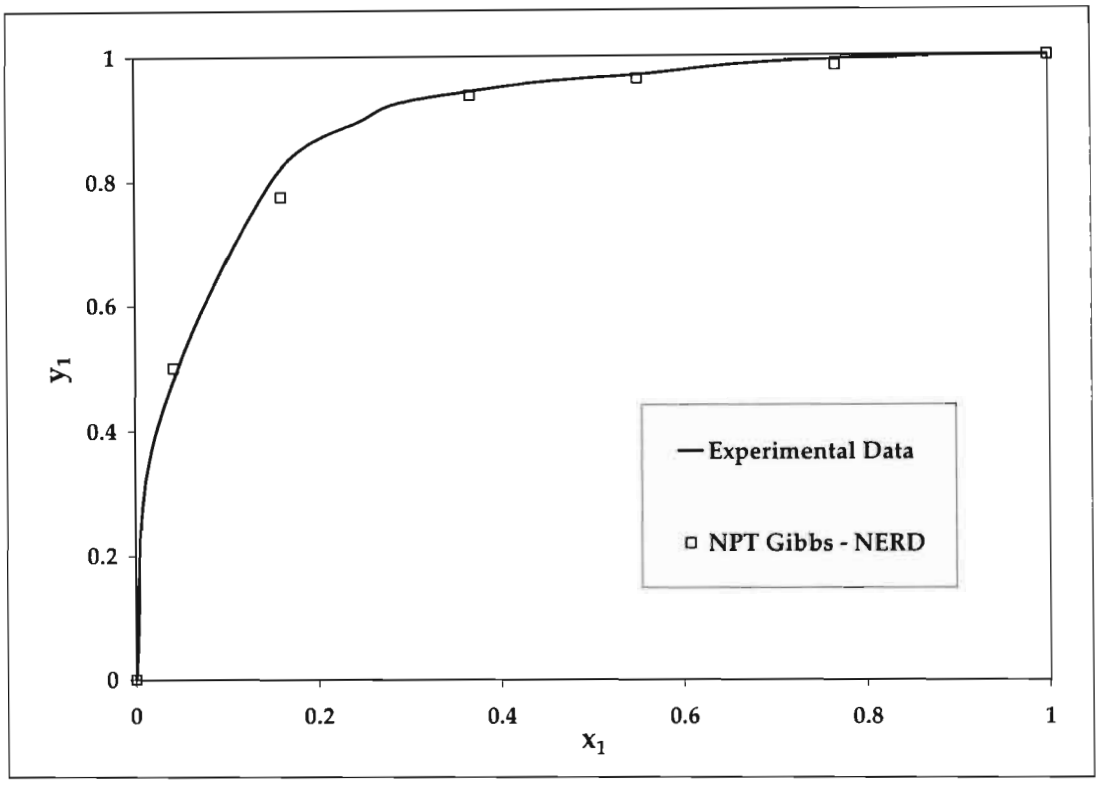

Figure 7.14: x-y plot n-pentane(1) + n-octane(2) mixture at 308.7K.

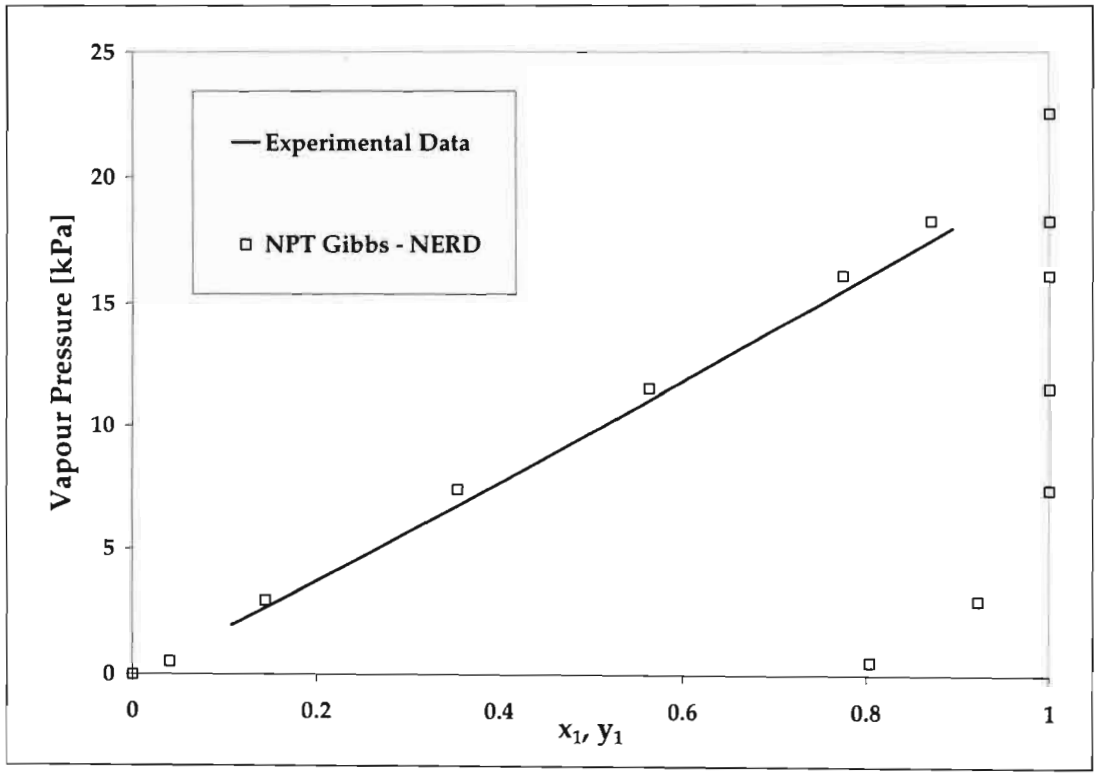

Figure 7.15: P-x-y plot for n-hexane(1) + n-hexadecane(2) mixture at 289.15K.

Chapter 8

Analysis and Discussion of Simulation Results

One may argue that all the work in this study is predictive (which, strictly speaking, is true) thus having a results chapter and an analysis and discussion chapter is inappropriate. However, for the purposes of this dissertation, the results (coexistence densities and vapour pressures) obtained directly from the simulations will be considered as results and any further properties determined from these data (such as critical properties) will fall under the analysis.

So far the required theory has been presented in Chapters 1 to 6 and the simulation results of this study were given in the previous chapter (Chapter 7). The focus of this chapter is to analyze and discuss these simulation results within the context of the available literature data, potential models applied and the simulation techniques and settings used during this study. Of particular interest will be the quality of

- the simulation results,
- the estimated critical properties,
- the three different potential models used.

It is important to note that the nature of this discussion will be largely qualitative. The many different aspects of the simulation methods and force field parameterizations have simply not been investigated sufficiently for one to approach the analysis of the simulation results in a highly quantitative manner. Thus, the exact effect of changing different

force field and simulation parameters cannot be accurately determined. Rather, the general trends and effects of these parameters must be noted. This is not to say that there will be no quantitative analysis, simply that the emphasis will be on the qualitative analysis.

8.1 Literature Data

Hydrocarbons are of great industrial importance (Ungerer, 2003). As a result, a large amount of effort has historically been put into investigating the properties of hydrocarbons. Unfortunately most of these studies only investigated the smaller hydrocarbon compounds. As a result, the available experimental data for larger hydrocarbon compounds are scarce and often contradictory (Siepmann et al., 1993). The focus of this study is not to compare MC simulation to the slew of alternative predictive methods, but rather to simply investigate MC simulation as an alternate method of determining vapour-liquid equilibria for long-chain hydrocarbons. Thus, for the purposes of this study, available experimental data has been considered paramount. The experimental data has been extrapolated (using the scaling law and the law of rectilinear diameters) in the situations where the data are limited. It is these data that the simulation results and predicted critical properties of this study have been compared with.

8.1.1 Experimental Data

As previously mentioned, much of the experimental work has been done for the smaller (shorter) hydrocarbon compounds.* The major difficulty in obtaining experimental data for the longer hydrocarbon compounds is due to the typical times required to equilibrate the long-chain hydrocarbon systems. Often these compounds are very costly and one is forced to use the same samples for many data points. Thus, the samples often spend up to 36 hours at elevated temperatures causing substantial thermal degradation (Harris, 2004).

The key sources of experimental data for this study were Smith and Srivastava (1986*a*), Smith and Srivastava (1986*b*), Vargaftik (1975), *DDB - Dortmund Data Bank* (2004), *NIST Chemistry WebBook* (2003) and *KDB - Korea Thermophysical Properties Data Bank* (2003). The data from the KDB has been considered the most questionable, and as such has only

*A quick investigation using the DDB software will show that there is very little data for hydrocarbons longer than dodecane.

been used when no other data could be found in any of the other sources (this was the case for 1-hexadecene, 1-eicosanol and the two binary systems investigated). The first three references are all data books of collected experimental work and they were all published prior to 1990! The most up to date source of experimental data used for this study was the DDB.

It was found that the phase coexistence data, in particular, was the most difficult to obtain. Often, in the cases where the data was available, the temperature ranges covered was often small and stopped well below the critical temperature. The scaling law and the law of rectilinear diameters were used to extrapolate these available experimental data sets to higher temperatures. The vapour pressure data was often available up to the normal boiling point, thereafter the data became as rare as the coexistence density data. When there was no available vapour pressure data they were predicted using using the Antoine equation using the parameter values obtained from the KDB. These predictions proved to be quite adequate for this study.

8.1.2 Previously Published Monte Carlo Simulation Data

There have been a large number of studies on hydrocarbons using MC simulations. For long-chain hydrocarbons the list is somewhat shorter, with only the n-alkanes (up to maximum size of C_{48}) being investigated to any extent by Siepmann et al. (1993), Smit et al. (1995) and Nath, Escobedo and de Pablo (1998). Typically, the focus has been on the smaller hydrocarbons for two main reasons. The first reason is that the short hydrocarbons are the ones used to parameterize the force fields. This is possible since all the intra- and intermolecular components of the potential energy are present once the hydrocarbon has a length five pseudo-atoms (see Chapter 4). The second reason is that the longer hydrocarbons take longer to simulate. There are more interactions per molecule that must be considered and more trial moves that are required for the CBMC method. Still, there was a fair body of MC simulation publications that were relevant to this study.

Almost all research in this field begins with investigations into n-alkanes. For this reason there is a large amount of published data, of which, the following were consulted the most during this study:

- SKS – Smit et al. (1995) and Smit et al. (1998),
- OPLS-UA – Jorgensen et al. (1984),

- TraPPE-UA – Martin and Siepmann (1998) and Martin and Siepmann (1999),
- NERD – Nath, Escobedo and de Pablo (1998) and Nath and de Pablo (2000),
- Buckingham Exp-6 – Errington and Panagiotopoulos (1998).

For the two n-alkane mixtures investigated Nath, Escobedo, de Pablo and Patramai (1998) was consulted. Substantially less work has been published on the 1-alkenes. Those consulted during this study were:

- TraPPE-UA – Wick et al. (2000),
- NERD – Nath et al. (2001*a*).

As was the case for the 1-alkenes, very little work has been published for the 1-alcohols. Even though there is an increase in the simulation complexity due to the charged nature of the alcohols, they are an industrially important hydrocarbon and are being investigated more and more. The following publications were consulted during this study:

- OPLS-UA – Jorgensen (1986),
- TraPPE-UA – Chen et al. (2001),
- NERD – Khare et al. (2004).

8.1.3 Simple Predictive Correlations Used

During this study, a number of simple predictive correlations were used to approximate the initial simulation conditions, extrapolate available experimental data and estimate the critical properties. The coexistence densities were mainly predicted using the Racket equation (Spencer and Danner, 1972) for the saturated liquid density and the Virial equation of state for the vapour density. The second Virial coefficients were predicted using the method of McGlashan and Potter (1962). Both the Racket and Virial equation of state methods require the critical properties and vapour pressures at the particular temperature. The critical properties were generally obtained from the DDB and the vapour pressures were estimated using the Antoine equation. These coexistence density methods were only used to approximate the initial simulation conditions required, the results were not directly used in any other way. It was found that these methods only performed adequately for the non-polar hydrocarbons shorter than ten pseudo-atoms. This

is largely because Racket was only fitted for molecules up to a carbon length of about 10 carbons, also the method of McGlashan and Potter (1962) was only for n-alkane (though it seemed to work adequately for 1-alkenes). For the alcohols and longer non-polar compounds a trial and error method of determining the initial condition was used.

The scaling law, law of rectilinear diameters were used to extrapolate available experimental coexistence density data to higher temperatures and estimate the critical properties presented later in this chapter. They are the standard method used for estimating the critical properties from simulation density results and are accurate to within the standard deviations of the Gibbs ensemble (Martin and Siepmann, 1998). The Clausius-Clapeyron equation was used to estimate the normal boiling point and the critical pressure. A key assumption in using it was that the heat of vaporization was constant over the temperature range considered. While this is strictly not true (Chase, 1987), for the relatively simple hydrocarbon compounds of this study these methods proved adequate.

A group contribution based equation of state method (Coniglio et al., 2000; Coniglio and Nouviaire, 2001) was investigated. The thermophysical properties of heavy hydrocarbons are predicted using the normal boiling point as the reference temperature. Unfortunately, this method could not be used during this study due to the fact that a number of the required volume correction factors had not been regressed yet. Still, the method shows promise and has since been simplified slightly by Crampon et al. (2004).

8.2 Simulation Details

There are a wide variety of MC simulation techniques that have been developed for different chemical systems. Excellent reviews of many of these techniques may be found in Allen and Tildesley (1987), Frenkel and Smit (2002) and Sadus (1999). The aim of this study was simply to investigate the ability of the standard MC simulation techniques in simulating long-chain hydrocarbons. The values used for the simulation settings have been clearly presented in the previous chapter. This section will focus on the the reasons for using particular techniques and parameter values, many of which are simply rule of thumb.

8.2.1 Gibbs Ensemble

The MC simulation method primarily used in during this study was the Gibbs ensemble method. While other methods were initially investigated, the Gibbs ensemble method proved the most appropriate due to its simple, direct (although discrete) nature. The Gibbs ensemble is a direct simulation method in that the vapour-liquid equilibria data is directly calculated during the simulation. Other methods, such as Histogram Reweighting, are more complex and require a number of further calculations before any practical vapour-liquid equilibria data can be obtained from the simulations. The Gibbs method is discrete - every simulation results in one vapour-liquid equilibrium point. This meant that the vapour-liquid coexistence curves had to be investigated point by point.

Deciding on values for the various setting related to the Gibbs ensemble method can be very difficult. There are an enormous amount of variables which are interdependent. The lack of transferability between different simulation situations also means that one is often forced to follow rules of thumb. Many of the setting used for the Gibbs method during this study were obtained from the following publications:

- Siepmann group from their TraPPE-UA series (Martin and Siepmann, 1998),
- Smit group from their work on CBMC methods (Vlugt, Martin, Smit, Siepmann and Krishna, 1998) and zeolites (Vlugt, Zhu, Kapteijn, Moulijn, Smit and Krishna, 1998).

Two other important research groups were also initially consulted. These groups follow slightly different paradigms from one another and from the above groups.

- Panagiotopoulos group for their Buckingham Exp-6 series (Errington and Panagiotopoulos, 1998),
- de Pablo group for their NERD series (Nath, Escobedo and de Pablo, 1998).

The two program codes used during this study were developed by the Smit (BIGMAC code) and Siepmann (TOWHEE code) groups. This is key reason for basing many of the simulation settings on their work. Both groups have been very influential in the MC simulation field. In fact it was the Smit and Siepmann groups that introduced the two key simulation techniques which made this study possible, viz. CBMC and DC-CBMC.

The Panagiotopoulos group has been no less influential, having introduced the Gibbs ensemble method (Panagiotopoulos, 1987; Panagiotopoulos et al., 1988). However, only very basic public codes have been released by the Panagiotopoulos group and none have been released by the de Pablo group. A number of specialized force field implementations and simulation techniques have been used by these two groups. For example the Buckingham Exp-6 potential model not only uses a different functional form (exp-6 form), the tail corrections are computed differently (McKnight, 2005).

In any practical application of MC simulation there are many settings that do not directly affect the results of the simulation. These settings affect the aesthetic aspects of the simulation, such as how often data is printed to screen or to an output file. These types of settings will not be discussed in this dissertation.

The Number Monte Carlo Cycles Used

The length Gibbs ensemble simulations are typically measured in terms of the number of MC cycles taken. For any given MC simulation, containing a fixed number of molecules, one MC cycle is defined as one set of attempted trial moves for every molecule in the system (i.e. cycles = moves/N). Any given MC simulation is split into two stages, the equilibration and the production stages.

The equilibration stage is the part of the Markov chain (simulation) during which equilibrium states are reached and sampled consistently. It is during this stage that a number of simulation setting are fine-tuned online. For example, the maximum volume and the maximum translational changes are optimized online every couple cycles. While this technically breaks the simulations reversibility, it still allows the system to reach equilibrated states. Fine-tuning in this manner is ideal since the states sampled during this stage of the simulation are not used to calculate the final results. It was stated in the previous chapter that a minimum of 30000 cycles were used for the equilibration stages. This truly is the minimum used during this study, and a number of the longer hydrocarbon compounds required up to 60000 more cycles. It is impossible to predict exactly how many cycles will be required to reach equilibrium for a particular simulation since the Markov chains produced are based on probabilities. Thus, one is forced to run simulations and check them periodically to see if equilibrium has been reached. The easiest way to determine whether a system has reached equilibrium or not is to plot the running averages of some of the important system variables, such as density (for chem-

ical equilibrium) and pressure (for mechanical equilibrium). Figure 8.1 shows a typical plot of the vapour density during the last few thousand cycles of equilibration stage of a simulation. One can see that the system is nearing equilibration. The scale of Figure 8.1

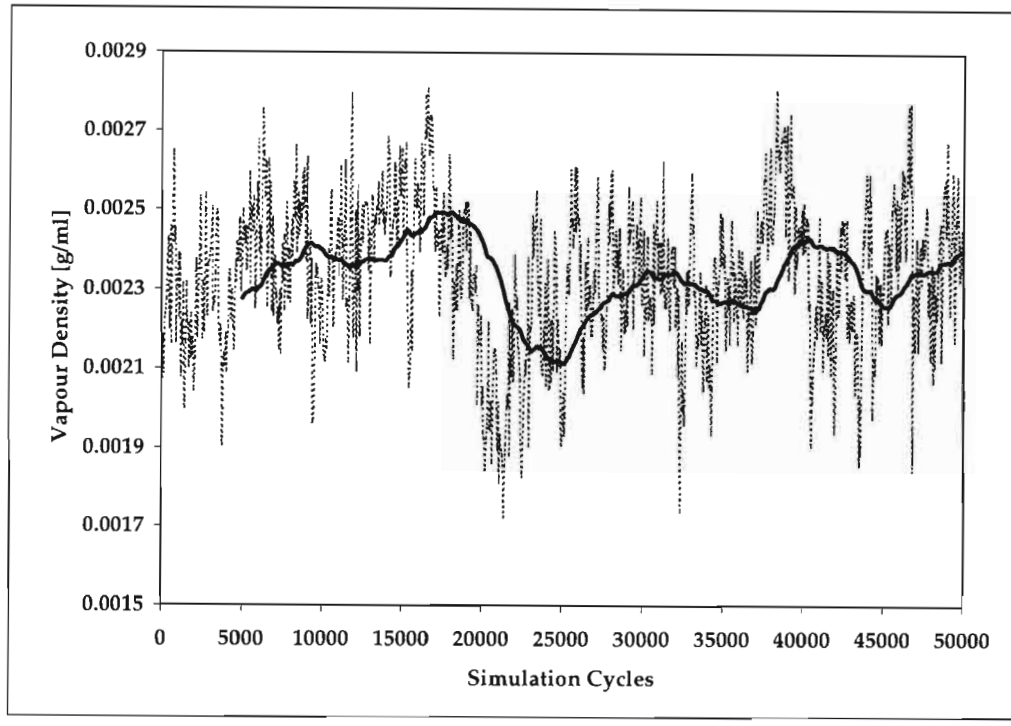

Figure 8.1: Plot of the vapour density during an equilibration stage of a simulation. The instantaneous values (dashed) and a running average (solid) are shown.

is important, although the system does not appear anywhere near equilibration, if one looks at the second half of the figure, the deviation of the running average is only about 10%. One must accept a certain standard deviation on any of the running averages, but how much of a standard deviation is dependent on the type of variable one is monitoring and the temperature of the simulation. Variables such as the vapour density and pressure will always have a greater standard deviation compared to the liquid density. This is largely due to the simulation method itself. The vapour phase typically consists of about 10–25% of the molecules in system, thus even a single molecule change affects the vapour phase more than the liquid phase. At higher temperatures the standard deviation that one must accept is higher than at lower temperatures. This is because the system is much more dynamic, many more extreme system configurations start being accepted. Once the running averages are effectively stable (not systematically increasing or decreasing) it may be assumed with confidence that equilibrium has been reached. This procedure of monitoring the running averages was the method used in this study.

BASH scripts were developed to effectively extract these data from the huge simulation output files. The basic BASH script used for extracting data is given in Appendix E. Appendix F shows a very abridged example of a TOWHEE output file. Once it was confirmed that equilibrium had been reached, the production stage was performed.

The production stage of the simulation is the stage during which all the thermophysical properties are determined. Before this stage may be started a number of the variables that were fine-tuned during the equilibration stage must be semi-fixed. These settings, such as the maximum volume and translational changes, were typically reduced to about 10 updates throughout the production stage of the simulation. This is important for maintaining the integrity of the Markov chain and speeding up the simulation during a stage that typically run for many hours. The number of cycles that the production stages were run for was between 50000 and 80000. The exact number of cycles was held constant for each system so as to insure consistency in the measurement of the standard deviations. It is assumed that the configurations sampled during the production stage are largely independent. This assumption allows one to split the production stage into 5–10 blocks and consider each part as an independent production stage. All the thermophysical properties were then calculated for each of these blocks and the average over all these blocks determined. In this manner the same final results were obtained but more representative standard deviations could be calculated. The number of cycles that one runs the production stage for depends on the properties being measured, the consistency of the results desired and the practicality of running the simulation for many hours. In this study, simulations were run until the standard deviation were typically about 1% for the liquid densities and about 10% for the vapour densities and pressures. Unfortunately, it was not always possible to maintain such strict restrictions on the standard deviations. The major difficulty was the fact that as the hydrocarbon length increased so more cycles were required to achieve the desired, low, standard deviations. For some systems the times required for simulating these long-chain hydrocarbons for many cycles became prohibitive. This was the case for a number of the 1-alkene simulations performed on Monolith (*National Supercomputer Center, Linköping University, Sweden*, 2003). The time limit on a single simulation meant that the production stages had to be split up. Thus, the final averages and standard deviation were calculated independently.

Simulation Times

The CPU time per cycle was not constant for the simulations. It was found to depend on many different factors. The factors that most affected the simulation times were:

- CPU speed,
- system size (number of interaction sites),
- CBMC and DC-CBMC parameters,
- potentials considered (both non-polar and polar),
- size of potential truncations,
- simulation temperature,
- types of trial moves attempted.

From this list of factors it is possible to see why it is so difficult to optimize simulation parameters for anything more than a very select set of situations. Instead, one is forced to follow rules of thumb and common sense when setting many of the simulation parameters. Many of these factors acutely affect the simulation of long-chain hydrocarbons.

There is little that one can do about the CPU speed, typically systems are only upgraded every 2–3 years. Rather than attempting to brute force the simulations, the better solution is often to carefully set the simulation parameters. Fortunately there is some freedom in specifying these simulation settings. These factors will be discussed throughout the remainder of this section.

Table 8.1 shows the approximate CPU times taken for the simulations performed during this study. The higher temperature simulation run-times were used so as to not over exaggerate the simulation times. Also, the number of cycles per point has been assumed to be 110000 cycles. This value was determined from the minimum number of cycles required (see the previous chapter). In many cases, particularly for the longer hydrocarbon compounds, at least 50000 extra cycles were required to achieve equilibrium. The number of points per system was determined by considering the number of force fields used times the number of points along the coexistence curve predicted using each force field.

Table 8.1 is not an exaggeration of the time required to simulate these systems! Clearly, this study would not have been possible without the use of the "Yoda Beowulf

Hydrocarbon	Min/Cycle	Time/Pnt [hrs]	Pnts/Sys	CPU Time/Sys [days]
n-octane	0.0153	28.03	8	9.34
n-hexadecane	0.0164	30.13	10	12.55
n-tetracosane	0.0235	43.10	8	14.37
1-octene	0.0169	30.98	12	15.49
1-decene	0.0179	32.83	10	13.68
1-hexadecene	0.0198	36.25	12	18.13
1-octanol	0.0230	42.17	15	26.35
1-dodecanol	0.0366	67.17	18	50.38
1-hexadecanol	0.0474	86.81	15	54.26
1-eicosanol	0.0664	121.70	18	91.28
		Total CPU Time Required [weeks]		43.69

Table 8.1: Approximate CPU times required for the long-chain hydrocarbons simulations.

Cluster" (Appendix B). From Table 8.1 it is clear that even minor speed-ups during each cycle can have massive time-saving benefits for these long-chain hydrocarbons. On many systems there are limits to the maximum time that these sort of runs can take. For example the Monolith system has a time limit of 100 hours per simulation run.

Setting the Fixed Probabilities

It is with the fixed probabilities that the distribution of the attempted trial moves are set. These values are kept constant over a whole simulation and are largely constant over whole systems. As stated in the previous chapter the following fixed probabilities were used:

- volume exchanges (0.006),
- molecule swaps (0.328),
- translational moves (0.222),
- rotational moves (0.222),

- conformational moves (0.222),

These values are based on Smit et al. (1995) and those typically used for the TraPPE-UA papers. The following rule of thumb method, used by the Siepmann group, was used in this study to set the fixed probabilities:

> To yield approximately one accepted molecule swap and/or volume exchange per 5–10 cycles; the remainder of the moves were split equally among the translational, rotational and conformational moves.

For the most part the values presented above were quite adequate, only for the longer hydrocarbon compounds were more molecule swaps required. It is of paramount importance for the Gibbs ensemble method that a reasonable number of molecule swaps (about 1 swap per 10 cycles) occur between the vapour and liquid phases. These swaps achieve the chemical equilibrium between the two phases. The volume exchanges achieve the mechanical equilibrium and the rest of the moves maintain the thermal equilibrium.

Summing both the volume exchange and molecule swap probabilities only amounts to about 0.334. Since the majority of the molecule swaps are not accepted and very few volume exchanges are performed, one can see that the translational, rotational and conformational moves are the most commonly attempted and accepted moves. The maximum translational and rotational changes are two parameters which are fine-tuned during the simulation to achieve a 50% acceptance rate. The conformational moves are molecule regrowths using the CBMC method and are typically split between complete and partial regrowths. These three moves settle the system down after the extreme configurational changes that occur when the volumes are exchanged or molecules are swaped.

Using Parallel Codes

Only serial computational codes have been used in this study. It is however important to mention that it is possible to perform parts of the MC simulation in parallel. The MC simulation method is in essence a sequential method. However, performing the attempted trial moves on different CPUs, then selecting the optimum trial move from these and continuing the simulation in this manner, one is able to perform parallel calculations for MC simulations. Unfortunately, new difficulties are introduced when using these parallel methods. There are programming difficulties associated with maintaining information integrity within a parallel code during the simulation and there is the

lag associated with the communication between the separate processes. For these reasons, parallel codes are typically only used when one is simulating very large systems (N $\sim$ 1000+) or very large molecules (e.g. polypeptides). These methods have been extensively discussed by Esselink et al. (1995), Loyens et al. (1995), Vlugt (2000) and McKnight (2005).

8.2.2 Configurational-bias Monte Carlo Methods

The use of CBMC methods during this study is possibly the single most important reason why these simulations were at all possible. Achieving a sufficient number of molecule swaps is vital to the Gibbs ensemble methods correct operation. Without the CBMC methods it would not have been possible to simulate hydrocarbons longer than a few pseudo-atoms in size. There is often little or no mention of the values used for the CBMC parameters in literature.[†] This is generally the case because the exact values depend largely on the molecules geometry, the temperature of simulation and the density of the system. Thus, in most cases the CBMC parameter values are not transferable. There are often differences in the actual implementation of the CBMC method and this clearly makes the exact parameter values non-transferable. For example, in Chen et al. (2001) the alcohol molecules were grown by first inserting the hydroxyl-O, then adding the hydroxyl-H and the α-carbon concurrently, and thereafter growing the rest of the molecule in a standard manner. However, these types of molecule insertions were not employed within this study. All molecules have been grown from one of the ends of the molecule.

Figure 8.2 shows the number of molecules during a simulation where the CBMC parameter values are too low and Figure 8.3 shows the number of molecules during a simulation where the CBMC parameters values are sufficient. In Figure 8.2 one can see that molecule swaps often do not occur for 50–100 cycles. As stated earlier this is way too few, the system will take too long to reach chemical equilibrium between the liquid and vapour phases. In this case the CBMC parameters have been set too low and must be be increased. Figure 8.3 shows a system where sufficient molecule swaps are occurring. The CBMC parameters used for this simulation have been set high enough. One must be careful not to set them too high else simulation time is waisted. When the swap moves are too low one must first make sure it is due to the CBMC parameters. This is done

[†]The TraPPE-UA series is the major exception to this statement.

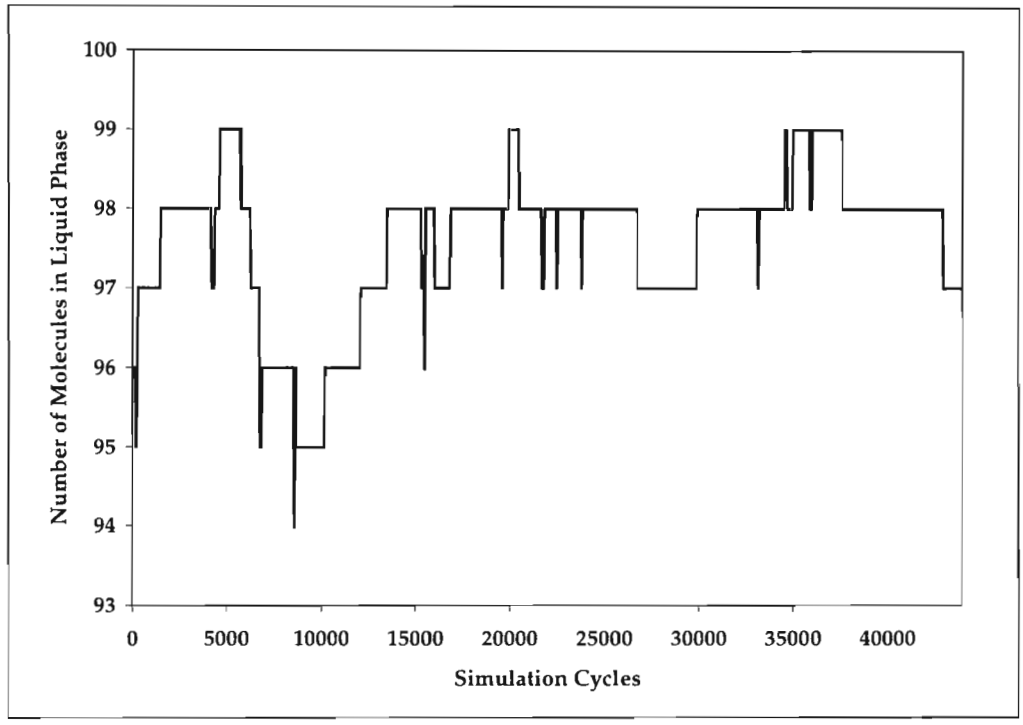

Figure 8.2: Plot of the number of molecules in the liquid phase during a simulation when the CBMC parameters have been set too low.

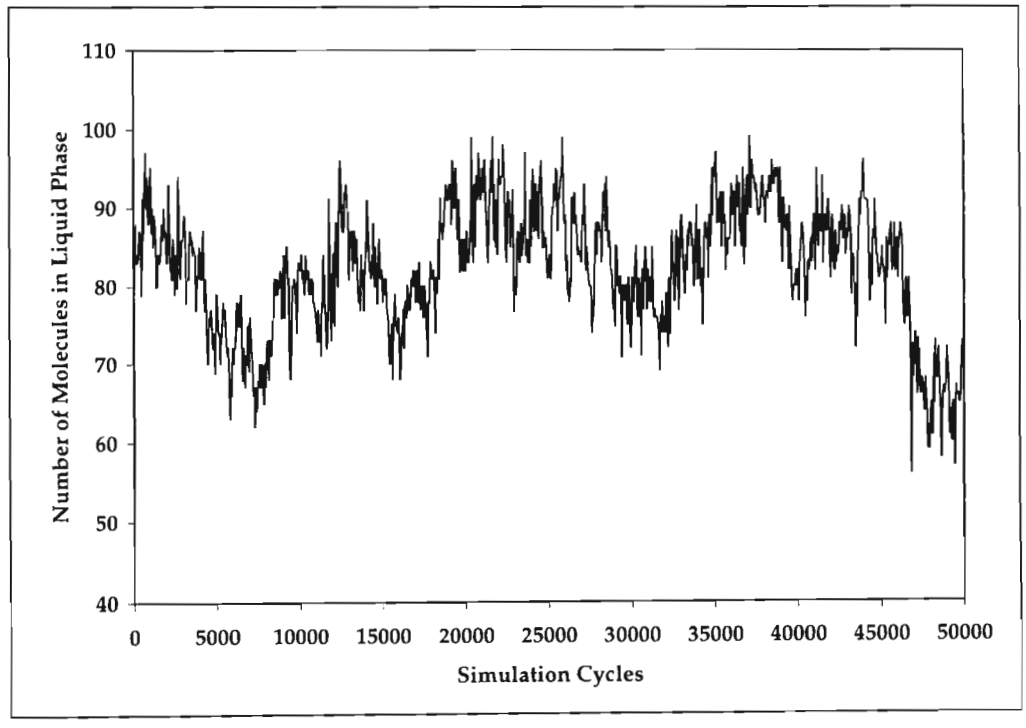

Figure 8.3: Plot of the number of molecules in the liquid phase during a simulation when the CBMC parameters have been set correctly.

by checking is whether the attempted insertions are being fully grown (this statistic is usually given with the number of accepted insertions at the end of a simulation). If most of the attempted insertions are being fully grown then it is likely that one will have to increase the number of attempted molecule swaps. On the other hand if the molecules are not being fully grown then the CBMC parameters must be increased. When the CBMC parameters are too low one must first decide which parameter values to increase and by how much. There are no standard procedures for solving these two issues so they must be resolved using simulation experience, trial and error and common sense.

Standard CBMC Settings

In the standard CBMC there are two parameters that must be optimally set, viz. the number of trial insertions for the first pseudo-atom, f, and the number of trial insertions for the subsequent pseudo-atoms, k. Studies have been performed to optimize values of f (Esselink et al., 1995) and k (Mooij and Frenkel, 1996). This standard CBMC was used with BIGMAC for the non-polar simulations and proved quite adequate. The geometrical simplicity of the compounds simulated (only linear chains) was an advantage because smaller values for f and k were possible and this sped up the simulations. For the C_8 chains the CBMC parameter values used were 10 for both f and k. However, it was found that as the chain length increased so the optimal values of f and k increased too. It was found by trial-and-error that an effective number of trials was about the number of pseudo-atoms in the hydrocarbon compound. For long-chain molecule one must be careful when setting these parameters since a change in f does not really affect the simulation times but even a small change in k can substantially slow down the simulations. This can be seen using a molecule with N_c pseudo-atoms. For this molecule there will be $f + k^{(N_c-1)}$ trials, thus the longer the chain-length of the molecule, the more impact an increase in k would have on the simulation time.

Coupled-decoupled CBMC Settings

For the coupled-decoupled CBMC there are many more parameters that must be set. The typical values that were used during this study have been presented in the Chapter 7. In Martin and Siepmann (1999) the standard values suggested are much higher than those used in this study. This was because the hydrocarbon compounds simulated in this study were geometrically simple (they are all linear) which limited the possible range of

bend and torsional angles.[‡] Still, as the chain-length increased so the effect of increasing the coupled-decoupled CBMC parameters increased dramatically. The number of trials as a function of the chain-length, N_c, is:

$$f + (n_{ext} \times n_{tor} + n_{bend} + n_{bond})^{N_c - 1} \tag{8.1}$$

The coupled-decoupled scheme used in this study was with the intermolecular and torsional potentials coupled and all the rest of the interactions decoupled. Thus, one had to be very careful not too over specify the values of these parameters. However, these values could be reduced still further since arbitrary trial distributions (Chapter 5) could be used with the TOWHEE code. In order to use the arbitrary trial distributions however, a number of extra parameters had to be set. These parameters were the standard deviations of the bend and torsional angles from their equilibrium values. Again the geometrical simplicity was an advantage since it limits the range of possible angles and thus lowered the standard deviations that could be set. These initial settings for 1-octanol using the arbitrary trial distribution method were suggested by Martin (2004*b*). For the other hydrocarbon compounds simulated the settings were checked (and modified if necessary) based on the trial simulations. During these trial simulations the arbitrary trial distributions were not used and the angles (plus bond length for NERD) were sampled. The file `towhee_movie` was created during the simulation and in this file run-time information about molecule positions and box sizes were stored. The standard deviations for the bond-length, bend and torsional angles were then determined using a code called `analyze_movie`. Based on these results the arbitrary trial distribution parameters were checked. It was found that even for the much longer hydrocarbon compounds the standard deviation values proved adequate. The values for the parameters shown in Equation 8.1 still had to be increased as the chain-length increased to maintain the number of molecule swaps.

Dual-cutoff CBMC Settings

This technique is used in combination with either of the above CBMC styles. It was also critical to the practical simulation of long-chain hydrocarbon compounds in this study. Tests performed early in this study showed that the use of DC-CBMC sped up

[‡]Of the three force fields, only NERD uses a variable bond length (which is still fairly stiff) so the geometry is not really affected.

simulations by a factor of at least two for C_8 and for the the longer hydrocarbons it sped them up by factors of as much as four. This method was originally developed by Vlugt, Martin, Smit, Siepmann and Krishna (1998) and has since been used with great success in many simulation situations, from adsorption of n-alkanes in zeolites (Vlugt, 2000) to more classical applications such as this study. It was shown in Chapter 5 that the optimum cut-off radius in DC-CBMC should be around 5Å (see Figure 5.6). This is the value used most often for studies investigating chain-like molecules and for this reason it was used during this study. This technique was particularly useful for the long-chain non-polar hydrocarbons since the only intermolecular forces were the van der Waals ones. This force decays rapidly ($\mathcal{U} \propto r^{-6}$) so truncating the potential around a newly inserted pseudo-atom proved very effective. Even for the polar long-chain hydrocarbons (1-alcohols), the method yielded excellent time savings since most of the molecule was still non-polar. In situations where the majority of the molecule was polar or different potential energy functional forms (Chapter 4) are used then it is highly advisable to re-calculate the optimum DC-CBMC cut-off radius. In fact, Martin (2004*a*) suggests that the cut-off radius be increased to 10Å when performing Coulomb simulations.

8.2.3 Potential Truncations

There are a number of aspects one must consider when setting the size of the potential truncation. Most importantly one must ensure that the radial distribution function is effectively unity (Chapter 5). Figure 8.4 shows the intermolecular radial distribution function for 1-octanol. The sharp peak near 2.5Å is due to hydrogen bonding occurring in the alcohol system. This peak would normally occur around 4Å (roughly the average diameter of 1-octanol) if hydrogen bonding was not present. One can see that this radial distribution function has effectively reached unity well before 14Å. Typically the radial distribution function reaches unity by about $3.5\sigma^*$ (where σ^* is the approximate molecule radius). For the longer hydrocarbon chains 14Å is less than $3.5\sigma^*$ but the cut-off still proved quite sufficient. Making the truncation too large slows the simulation down and also sets a lower limit to the system sizes that may be simulated. During this study the Lennard-Jones potential was truncated using a spherical potential truncation of 14Å. This is the standard cut-off used in the TraPPE-UA force field. The other two force fields, NERD and OPLS-UA, both use standard cut-offs which are shorter than this. It is important to know the cut-off used for a particular force field when it was parameterized.

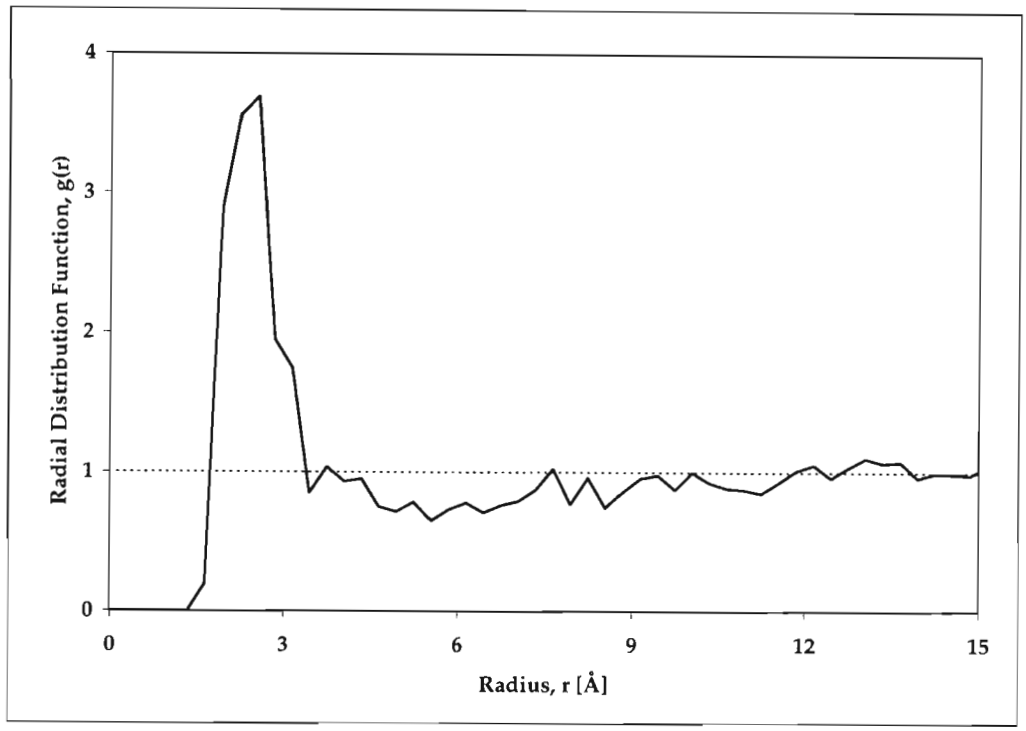

Figure 8.4: Intermolecular radial distribution function for 1-octanol.

Any subsequent use of that force field should not use a cut-off shorter than its original one since this may well bias the results slightly. This can be seen with the SKS force field (Smit et al., 1995). Here the authors did not include a factor of σ^3 in the tail correction which amounted to them effectively not using a tail correction. It was later shown, once the error was corrected (Smit et al., 1998), that the effect of now including the tail correction was a 4–6% increase in the predicted thermophysical properties.

8.2.4 System Sizes

In any simulation the system size selection is a very important aspect. The system size affects how long the simulation is going to run for and the accuracy of the results produced. Clearly, the larger the system size the more accurate the results but the longer the simulation will take. For long-chain hydrocarbons the effect is even greater. Consider a system of N molecules of chain-length N_c, the number of interaction sites would be $N \times N_c$. One can see that the systems quickly become impractical for many long-chain molecules to be simulated. However, there are limits to how small the system may be chosen as well. If too few molecules are used in the simulations then finite size effects cause the results to be biased. The potential truncation also limits how small the system

size may be. Rigorous finite size testing was not possible for these long-chain hydrocarbons due to the simulations times that were required to simulate the very large systems. The system sizes for hydrocarbon compounds up C_8 have already been simulated and published (Martin and Siepmann, 1998; Wick et al., 2000; Chen et al., 2001). The publications typically used 100–150 molecules for the C_8 systems and these same system sizes were used for the longer hydrocarbon chains. Since Siepmann et al. (1993) performed simulations using about 100 molecules for n-alkanes up to a length of 48 pseudo-atoms the use of these system sizes appeared valid. Once the number of molecules for the simulation had been decided, the simulation box sizes had to be set. Typically, the initial liquid box size was adjusted to be between 31Å and 35Å, so that the vapour box had at least 10 molecules in it. In setting the box sizes one had to ensure that the liquid box size was not set too small. This was due to the constraints of minimum image and the potential truncation (set to 14Å). In Figure 8.5 the change in the linear box length during a simulation has been shown. From this figure one can see that the linear box length

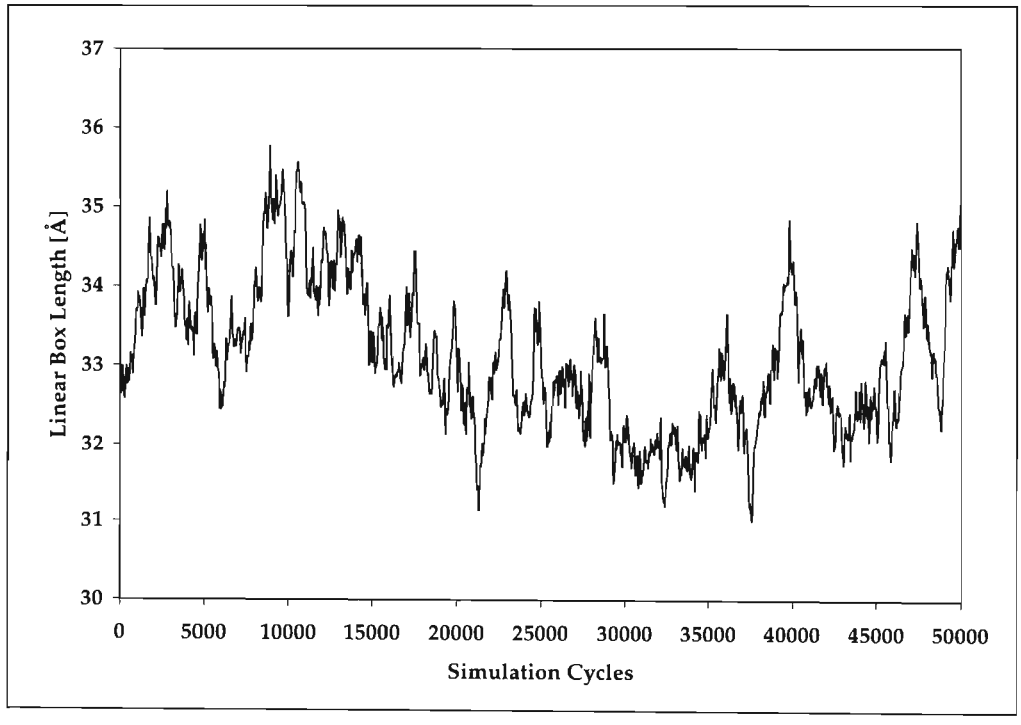

Figure 8.5: Plot of the linear box length during a simulation.

varies by about 15%. Minimum image forces the box size to be at least twice the potential truncation (i.e. linear box length > 28Å). Thus, linear box lengths smaller than about 31Å were avoided since random fluctuations in the box sizes could cause this minimum

to be reached. If the box size did go below this then the simulation effectively failed and had to be restarted. In some cases if this only occurred during the early parts of the equilibration stage then the simulations could be continued. However, in most cases the simulation had to be restarted using larger initial box lengths. For the longer hydrocarbon chains the initial liquid box lengths often were about 35Å since the liquid density of these compounds were so high.

8.2.5 Ewald Summation Settings

The Ewald summation technique was only used for the 1-alcohol simulations performed with the TOWHEE code. It was used to determine the Coulomb potential energy contribution. The Ewald summation was needed since the Coulomb potential is a long-ranged force ($u_{Coul} \propto r^{-1}$) and could not be truncated in the same manner as the van der Waals potential. The sharp point charges associated with the α-carbon, hydroxyl-O and hydroxyl-H are screened using a gaussian charge distribution. This enables the screened energy to be determined and then corrected for. The method has been presented and discussed in Chapter 5.4.The tin-foil boundary condition was used with $\kappa \times L = 5$ and $k_{max} = 5$. The 'tin-foil boundary condition' amounts to a neutralizing system boundary or $\epsilon_s = \infty$. The values used in this study are directly from Chen et al. (2001). They are effectively those suggested by Allen and Tildesley (1987) (shown in Chapter 5) and are the typical values used for simulations involving hydrocarbon compounds. While this method is considered the standard method of dealing with the Coulomb potential energy, there other methods which are faster for larger systems. It is shown in Frenkel and Smit (2002) that the Ewald summation scales with $\mathcal{O}(N^{3/2})$ whereas these other methods scale with $\mathcal{O}(N \log(N))$ (particle-particle/particle mesh method) and $\mathcal{O}(N)$ (fast multipole method). Due to time limitations these alternative methods were not investigated during this study and further details on these methods may be found in Frenkel and Smit (2002). Khare et al. (2004) made use of the charge-group based cutoff approach rather than the Ewald summation because it was claimed that the charge-group based cutoff approach is more efficient and comparable in accuracy to the Ewald summation method. However, there was very little literature available which made use of this method so the Ewald summation was used for the NERD simulations as well.

8.3 Potential Models Investigated

Three potential models were investigated as part of this study: NERD, OPLS-UA and TraPPE-UA. All three of these force fields use the Lennard-Jones 12-6 functional form to describe the non-bonded van der Waals interactions. The intramolecular interactions have all been described using a similar set of functions. All the details of these force fields have been presented in Chapter 4. Over the past two decades many force fields have been developed, some of these force fields use similar functional forms as those of NERD, OPLS-UA and TraPPE-UA but many force fields also use very different functional forms. In most cases the force fields have been developed to simulate systems under particular conditions. For example, the OPLS-UA was parameterized specifically to perform liquid simulations at 25°C. The force fields used in this study were chosen for a number of reasons. The key reason for choosing the NERD and TraPPE-UA force fields was that they were specifically parameterized to perform vapour-liquid simulations for hydrocarbon compounds and the authors of these force fields are steadily adding newly parameterized pseudo-atoms. The OPLS-UA force field is considered the historical standard force field due to its early and effective parameterization. It was chosen for the 1-alcohol simulations since no previous simulations could be found in literature for any 1-alcohols longer than 1-octanol for any force field. All the force fields that were selected use the same functional forms because Chen et al. (1998) showed that one cannot directly compare force fields when the functional forms were different. This is due to the fact that different functional forms enforce different assumptions upon the nature the repulsive and attractive forces. All of the three force fields used in this study are united-atom force fields, i.e. it was assumed that the hydrogens could be collapsed into their carbon centers. This made the simulations practical since it reduced the number of interaction sites substantially. In fact, the increase in computational time would have been prohibitive if all-atom simulations had been attempted. Consider a system of N n-alkane molecules each with a carbon number of N_c. For an united-atom system there are $N \times N_c$ interaction sites, but for an all-atom system there are $N(3N_c + 2)$. This number of interaction sites only considers the non-polar forces. However, most all-atom force fields also use Coulomb interactions between each pseudo-atom. By considering the polar interactions as well, the number of interaction sites for the all-atom system would double. An interesting alternative to carbon-centered united-atom force fields were the

anisotropic united-atom force fields. Unfortunately, simulations using these force fields were beyond the scope of this study.

8.3.1 A Brief Comparison Between the Force Fields

The major difference between these three force fields are the intention behind their use. The TraPPE-UA force field (Martin and Siepmann, 1998) continued attempts, begun in the 1990's with SKS (Siepmann et al., 1993; Smit et al., 1995), to create a simple, united-atom force field specifically for hydrocarbon compounds. It has since been shown that SKS was incorrectly parameterized due to an error in the code (Smit et al., 1998). Shortly after TraPPE-UA was introduced, the NERD force field (Nath, Escobedo and de Pablo, 1998) was introduced. The focus of TraPPE-UA was to minimize the number parameters required to yield reasonable predictions of thermophysical properties. However, the focus of the NERD force field was to give accurate predictions of thermophysical properties. This was partially achieved by introducing a large number of different pseudo-atoms, particularly for the shorter hydrocarbons (smaller than C_4). NERD also included bond-stretching whereas TraPPE-UA and OPLS-UA did not, although, this bond-stretching is fairly stiff due to it's high k_l/k_B. OPLS-UA, one of the oldest force fields still used today, runs the middle ground of these two ideals. It uses more pseudo-atoms than TraPPE-UA for the very short molecules but not as many as NERD. The bond-angles were fixed for OPLS-UA, however, most modern force fields include bond bending. The intramolecular interactions used for NERD and TraPPE-UA are derived from those used in OPLS-UA and are therefore fairly similar, both in functional form and parameter values.

Table 8.2 shows the hydroxyl functional group parameters for NERD, OPLS-UA and TraPPE-UA. One can see that all three force fields use the same LJ parameters for the α-carbons as was used for the non-polar CH_2 pseudo-atom, but with a positive charge assigned to it. This assumption, like the assumption that the CH_2 pseudo-atoms are transferable, helps to simplify the force fields substantially. The most interesting difference between the three force fields is the use of a ghost pseudo-atom (no LJ parameters) by OPLS-UA and TraPPE-UA and real pseudo-atom by NERD for the hydroxyl H. Chen et al. (2001) elected to use a ghost pseudo-atom with a charge for TraPPE-UA following many of the arguments that were used for OPLS-UA. The main reason was that the hydroxyl H was effectively a floating charge since the H was so small. However, Khare

Force Field	Pseudo-atom	ϵ/k_B [K]	σ [Å]	Charge
NERD	α-CH_2	45.8	3.93	+0.29
	O	108	2.98	-0.71
	H	3.89	0.98	+0.42
OPLS-UA	α-CH_2	59.38	3.905	+0.265
	O	85.5	3.07	-0.70
	H	0	0	+0.435
TraPPE-UA	α-CH_2	46	3.905	+0.265
	O	93	3.02	-0.7
	H	0	0	+0.435

Table 8.2: The -OH alcohol functional group parameters.

et al. (2004) opted to use a hydroxyl H with small LJ parameters. Khare et al. (2004) used the same argument as Nath (2003) used when parameterizing H_2S: that it added more flexibility to the model. This difference makes little difference to the predictions made by NERD compared with those of TraPPE-UA for these long-chain 1-alcohols, except for possibly fine-tuning the predictions.

Interestingly, both NERD and TraPPE-UA use only one torsional potential to describe the energy for CH_3-CH_2-CH_2-CH_2- and -CH_2-CH_2-CH_2-CH_2- whereas OPLS-UA uses different ones. Tests done during this study showed that this assumption didn't affect the simulations results substantially. However, as one makes more and more of these assumptions, slowly the results become lese and less accurate. As with many aspects of molecular simulations there is the difficult balance between transferability, complexity or speed and the accuracy of the predictions. Generally it is the consistency of the parameters within a particular force field that is most important for producing reasonable simulation results.

8.3.2 Effect of the Different Pseudo-Atoms

When one considers long-chain hydrocarbons, the relative effect of the different pseudo-atoms must be considered. There are a number of pseudo-atoms that have been used to simulate these hydrocarbon compounds but the accuracy of some pseudo-atoms are more important than others. Figure 8.6 shows a plot of the experimental normal boil-

ing points and critical temperatures as a function of chain-length. A similar plot of the

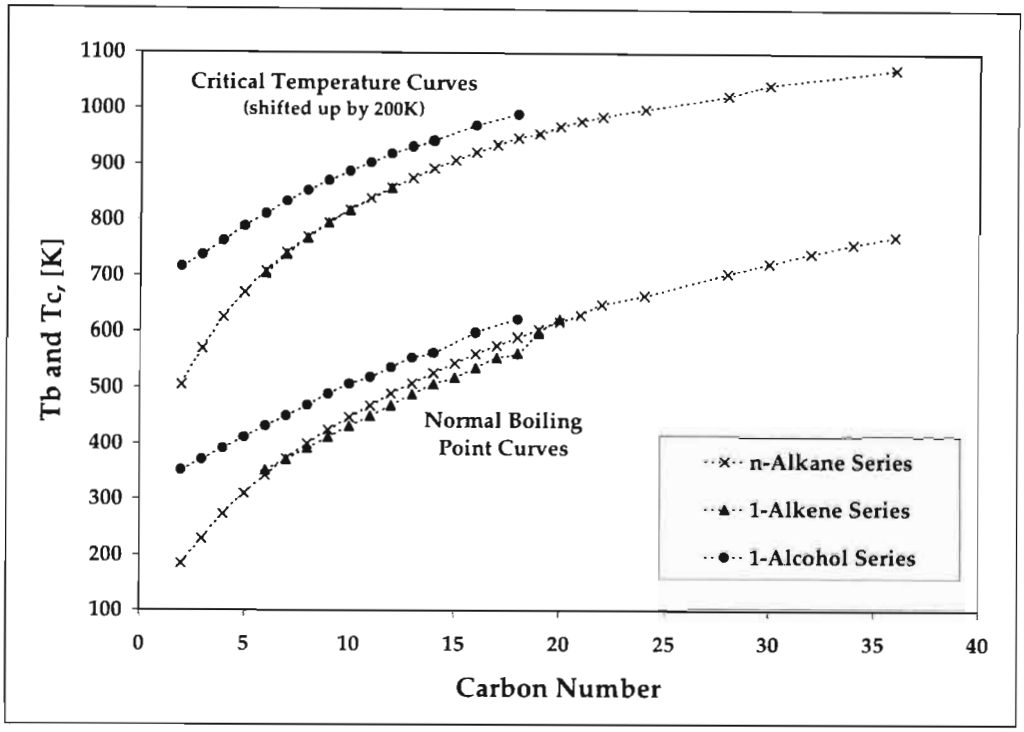

Figure 8.6: Plot of the normal boiling points and critical temperatures as a function of chain-length. All the data shown is experimental data obtained from the DDB.

critical density as a function of chain-length has not been shown since the critical densities are similar for any particular chain-length. From Figure 8.6 one can see that as the chain-length increases so the normal boiling points and critical temperatures tend toward the n-alkane values. If one considers a n-alkane molecule, as the chain-length increases so the effect that the CH_3 pseudo-atoms have on the thermophysical properties compared to the CH_2 pseudo-atoms decreases. In fact, for n-octane the contributions to the intermolecular potential due to the CH_3–CH_2 interactions are almost equal to the CH_2–CH_2 interactions. However, for n-alkanes longer than this the CH_2–CH_2 intermolecular interactions begin to dominate, increasing linearly, until for n-tetracosane the CH_2–CH_2 interactions contribute about four times as much as the CH_3–CH_2 interactions. Thus, the long-chain n-alkane effectively becomes a chain of CH_2 pseudo-atoms. By this one can see that it is very important for the CH_2 pseudo-atom, for any particular force field, to be as accurately represented as possible. For NERD and TraPPE-UA the LJ-well depths (ϵ/k_B) are effectively the same, 46K versus 45.8K, but the LJ-sizes (σ), 3.75Å versus 3.91Å, are slightly different. Thus, it is reasonable that the long-chain simulation results for NERD and TraPPE-UA are fairly similar. The CH_2 pseudo-atom for OPLS-

UA has a very different LJ-well depth compared to NERD and TraPPE-UA (59.38K) so one should explain some of the large differences between the predictions of OPLS-UA and the other two force fields. From this discussion, it seems likely that the functional groups associated with the 1-alkenes (CH_2=CH-) should effect the thermophysical properties less as the chain-length increases. Since the 1-alkene functional group is non-polar the effect should not be that dramatic either. In Figure 8.6 one can see that this is the case, even for the shorter 1-alkenes shown. The Scaling/Rectilinear curves used for the 1-alkenes (Wick et al., 2000) use the same critical component, $\gamma = 0.36$, thus the general shape of the coexistence curves are also similar in shape. The 1-alcohols have a polar hydroxyl functional group. It is logical to assume that the thermophysical properties should be affected by this group for even relatively long 1-alcohols. From Figure 8.6 it seems that the properties should become similar from chain-lengths of about 20–25 carbons. Thus, it may be possible to simulate very long hydrocarbon chains using only CH_2 pseudo-atoms. While simulations of this type may result in realistic predictions for the normal boiling points and critical properties, it is not likely to display other affects such as aggregation.

8.4 Estimation of the T_b's and Critical Properties

The vapour-liquid coexistence curves presented in the previous chapter were not the only objective of this study. The other key objective was to estimate the normal boiling points and the critical properties for the long-chain hydrocarbon compounds investigated. These properties were estimated by regressing the vapour-liquid coexistence simulation results.

8.4.1 Regression Method

The normal boiling points were obtained by interpolating the vapour pressure results using Clausius-Clapeyron plots. The critical properties were obtained from least-squares fits of the saturated liquid and vapour densities to the scaling law and the law of rectilinear diameters. These equations have been described in Chapter 6 and do not take finite size effects into account. Table 8.3 shows the typical least-squares regression used to estimate the normal boiling points and critical properties.

The factors '10' and '1000' in columns '2' and '4' of Table 8.3 were used to scale the

T [K]	$10(\rho_L - \rho_V)^{1/\gamma}$	T-T_c	$1000(\rho_L + \rho_V)/2$	1000/T	ln(P)
500	2.4845	-189.72	336.91	2.0000	4.1695
530	2.0717	-159.72	323.01	1.8868	4.9344
560	1.6985	-129.72	311.12	1.7857	5.6044
590	1.2683	-99.72	293.65	1.6949	6.0463
620	0.9202	-69.72	284.89	1.6129	6.5712
650	0.5253	-39.72	267.19	1.5385	6.9675
m_1	-0.0130	m_2	-0.4576	m_3	-6.0246
y-int_1	8.9865	y-int_2	250.3024	y-int_3	16.2775
r_1^2	0.9994	r_2^2	0.9950	r_3^2	0.9975

Table 8.3: Least-squares regression TraPPE-UA simulation data for 1-dodedcanol. The value of γ used was 0.29.

values to more similar sizes. These factors were taken into account in Equations 8.2 to 8.9. This set of equations was used to determine the normal boiling points and critical properties from the regressed data presented in Table 8.3:

$$\mathcal{A} = \left(\frac{m_1}{10}\right)^{\gamma} \tag{8.2}$$

$$\mathcal{B} = \frac{m_2}{1000} \tag{8.3}$$

$$C = \text{y-int}_3 \tag{8.4}$$

$$C' = 1000 m_3 \tag{8.5}$$

$$T_c = \frac{\text{y-int}_1}{|m_1|} \tag{8.6}$$

$$\rho_c = \frac{\text{y-int}_2}{1000} \tag{8.7}$$

$$P_c = \exp\left[C + \frac{C'}{T_c}\right] \tag{8.8}$$

$$T_b = \frac{C'}{\ln(101.325) - C} \tag{8.9}$$

The intermediate regression variables $\mathcal{A}$, $\mathcal{B}$, C, C′ have not been presented in this dissertation since they were only used to determine the critical properties (or calculated as a by-product of the procedure). The values of these intermediate regression variables were not required for anything else in this study. Table 8.4 shows the results of using Equations 8.2 to 8.9 on the data presented in Table 8.3.

Regressed	Parameters	Critical	Properties
$\mathcal{A}^{1/\gamma}$	-1.303×10^{-3}	T_c [K]	690
$\mathcal{B}$	-4.576×10^{-4}	ρ_c [g/ml]	0.2503
C	1.628×10^{1}	P_c [kPa]	1887
C′	-6.025×10^{3}	T_b [K]	517

Table 8.4: Results of the least-squares regression from Table 8.3. The value of γ used was 0.29.

These regression methods do not take into account finite size effects. As stated earlier (Section 8.2.4), methods taking finite size effects into account would have required simulations over a range of system sizes. This procedure would have taken prohibitively long for the long-chain hydrocarbon compounds dealt with in this study. It was shown by Martin and Siepmann (1998) that the error introduced by using the above regression method was about 1%, which is within the statistical deviations associated with Gibbs ensemble simulations. In previous publications it has been shown that these regression methods are adequate:

- TraPPE-UA (Martin and Siepmann, 1998; Wick et al., 2000; Chen et al., 2001)
- NERD (Nath, Escobedo and de Pablo, 1998; Nath et al., 2001*a*; Khare et al., 2004)

From these publications, values for the critical component, γ (see Equation 6.15), were also obtained. For the n-alkane and 1-alkene hydrocarbon compounds a value of 0.36 was used. This value is considered standard for non-polar hydrocarbon compounds. For the 1-alcohols a value of 0.29 was used. The reason that γ is lower is because the vapour-liquid coexistence curves for the alcohols has a different shape compared to the n-alkanes and 1-alkenes. Chen et al. (2001) was able to show that the corrected critical component value for alcohols ranged from 0.28 for the shortest alcohols up to 0.29 for the longer ones. This value of 0.29 was used by Khare et al. (2004) for the NERD force field as well.

8.4.2 Estimated Critical Properties

In this section the estimated normal boiling points and critical properties have been presented for each of the hydrocarbon groups studied. For the n-alkanes - Tables 8.5, the 1-alkenes - Table 8.6 and the 1-alcohols - Table 8.7.

n-Alkane	Force Field	T_b [K]	T_c [K]	ρ_c [g/ml]	P_c [kPa]
n-octane	NERD	394	576	0.2308	2928
(C_8H_{18})	TraPPE-UA	388	571	0.2382	2948
	Exp (DDB)	398	568	0.2322	2490
n-hexadecane	NERD	551	721	0.2211	2075
($C_{16}H_{34}$)	TraPPE-UA	545	722	0.2294	1884
	Exp (DDB)	560	725	0.2198	1400
n-tetracosane	NERD	651	802	0.2126	1246
($C_{24}H_{50}$)	TraPPE-UA	657	818	0.2124	1177
	Exp (DDB)	664	800	0.2117	870

Table 8.5: Comparison of the interpolated normal boiling points, critical temperatures, critical densities and critical pressures for selected n-alkanes

For the n-alkanes (Table 8.5) one can see that the normal boiling points are estimated accurately to within 2% (NERD) and 3% (TraPPE-UA) of the experimental values. The critical temperatures are estimated to within 2% (both NERD and TraPPE-UA) of the experimental values. The critical densities were estimated to within 1% (NERD) and 4% (TraPPE-UA) of experimental data. Both TraPPE-UA and NERD performed badly in the prediction of the critical pressure, 48% (NERD) and 35% (TraPPE-UA) compared with experimental data.

1-Alkene	Force Field	T_b [K]	T_c [K]	ρ_c [g/ml]	P_c [kPa]
1-octene	NERD	394	567	0.2398	2680
(C_8H_{16})	TraPPE-UA	383	565	0.2439	2900
	Exp (DDB)	394	567	0.2398	2680
1-decene	NERD	437	619	0.2336	2710
($C_{10}H_{20}$)	TraPPE-UA	435	620	0.2407	2928
	Exp (DDB)	443	617	0.2402	2220
1-hexadecene	NERD	547	711	0.2239	1545
($C_{16}H_{32}$)	TraPPE-UA	549	718	0.2192	1456
	Exp (KDB)	558	717	0.2459	1330

Table 8.6: Comparison of the interpolated normal boiling points, critical temperatures, critical densities and critical pressures for selected 1-alkenes

Table 8.6 shows the critical property estimates for the 1-alkenes. Here, the normal boiling points were estimated to within 2% (NERD) and 3% (TraPPE-UA) of experimental data. For the critical temperatures both NERD and TraPPE-UA estimated the experimental values to within 1%. The estimated critical densities varied largely from the experimental values. While the shorter chains were estimated to within 3% (NERD) and 2% (TraPPE-UA), for 1-hexadecene the estimates were only within 9% (NERD) and 11% (TraPPE-UA) of experimental data. Both NERD and TraPPE-UA performed much better for the 1-alkenes compared with the n-alkanes for the estimation of the critical pressures, being within about 20% (NERD) and about 30% (TraPPE-UA) of the experimental data.

1-Alcohol	Force Field	T_b [K]	T_c [K]	ρ_c [g/ml]	P_c [kPa]
1-octanol	NERD	463	639	0.2685	2959
($C_8H_{17}OH$)	OPLS-UA	510	714	0.3075	5140
	TraPPE-UA	460	627	0.2680	3015
	Exp (DDB)	468	652	0.2570	2777
1-dodecanol	NERD	510	705	0.2561	2056
($C_{12}H_{25}OH$)	OPLS-UA	599	802	0.3045	1575
	TraPPE-UA	516	689	0.2503	1887
	Exp (DDB)	535	719	0.2595	1994
1-hexadecanol	NERD	567	742	0.2542	1629
($C_{16}H_{33}OH$)	OPLS-UA	655	870	0.3101	2847
	TraPPE-UA	578	728	0.2699	1658
	Exp (DDB)	598	770	0.2551	1460
1-eicosanol	NERD	615	777	0.2281	981
($C_{20}H_{41}OH$)	OPLS-UA	750	932	0.3034	738
	TraPPE-UA	612	771	0.2524	1413
	Exp (KDB)	629	809	0.2542	1300

Table 8.7: Comparison of the interpolated normal boiling points, critical temperatures, critical densities and critical pressures for selected 1-alcohols

The final hydrocarbon group investigated was the 1-alcohols (Table 8.7). All three force fields were used to simulate these hydrocarbon compounds. It was found that the NERD and TraPPE-UA force fields were far superior to OPLS-UA for the prediction of normal boiling points and critical properties. Even though both NERD and TraPPE-UA

proved superior to OPLS-UA, their estimations were still the worst out of all three hydrocarbon groups. The normal boiling point estimations were within 5% (NERD), 4% (TraPPE-UA) and 20% OPLS-UA of experimental data. For the critical temperatures the estimates were within 4% (NERD), 5% (TraPPE-UA) and 15% (OPLS-UA) of experimental data. Critical densities were estimated to within 10% (NERD), 6% (TraPPE-UA) and 22% (OPLS-UA) of experimental data. Surprisingly, the estimated critical pressures were generally better than both the n-alkanes and 1-alkenes (for NERD and TraPPE-UA) estimations. Critical pressures were estimated to within 25% (NERD), 14% (TraPPE-UA) and 95% (OPLS-UA) of experimental data.

8.5 Simulation Results

The key objective of this study was to determine the vapour-liquid coexistence curves and critical properties of long-chain (longer than C_8) hydrocarbon compounds using MC simulation. Many important aspects of MC simulations have been discussed in this chapter. All these aspects have been considered and used to produce the vapour-liquid coexistence data presented in the previous chapter and this chapter. The focus of this section is to discuss these results.

8.5.1 n-Alkanes

The n-alkanes were the first group of hydrocarbons investigated in this study. Their geometrical simplicity and lack of partial charges make them an ideal entry point into investigations of long-chain hydrocarbons. Three hydrocarbon compounds were investigated, n-octane, n-hexadecane and n-tetracosane. Only the NERD and TraPPE-UA force fields were used to simulate these compounds since it has already been established by Nath, Escobedo and de Pablo (1998) that the OPLS-UA force fields was inferior in predicting alkane thermophysical properties. While there was experimental data available for n-octane, there was very little for n-hexadecane and even less for n-tetracosane.

Vapour-Liquid Coexistence Curves

The vapour-liquid coexistence curves for the n-alkanes have been shown in Figure 7.1. One can see that both NERD and TraPPE-UA predict the entire shape of the coexistence curves rather well. Typically the TraPPE-UA force field over-predicted the liquid and

vapour densities, whereas NERD tended to under-predict the liquid density and over-predict the vapour density. The force field parameters for NERD and TraPPE-UA are very similar for the n-alkanes. The significant difference between NERD and TraPPE-UA are the LJ-sizes (σ) for the CH_3 and CH_2 pseudo-atoms. The NERD LJ-sizes are about 5% larger than the TraPPE-UA LJ-sizes. This means that the NERD pseudo-atoms fill more space than they should and thus fewer pseudo-atoms are required to maintain the density and therefore NERD under-predicts the liquid density. The reverse happens for the TraPPE-UA pseudo-atoms and thus TraPPE-UA slightly over-predicts the liquid density. At the low densities encountered in the vapour phase this argument may not be used. Unfortunately, since both force fields over predict the vapour density, it is impossible to determine the exact cause of the over-prediction from these results. As the chain-length of the n-alkanes increases there appears to be a small increase in the difference between liquid density predictions of NERD and TraPPE-UA. However, there is no such increase in the differences for the vapour density. This trend is also due to the difference in the LJ-sizes since as the chain-length increases so the effect of the LJ-size increases. While these trends do not appear to be temperature dependent there is an increase in the standard deviations of the densities as the temperature increases. However, it was shown earlier in this chapter that the standard deviations do increase as the simulation temperature increases due to the nature of MC simulations. Both NERD and TraPPE-UA were fitted to saturated liquid density data and critical temperatures. This is a possible reason why the vapour density was consistently over-predicted by both NERD and TraPPE-UA. By not using the vapour density data in the parameterization of the force fields a bias was introduced for the vapour density. More recent potential models have been parameterized using more than just liquid density data and critical temperatures, although in order to do so they have had to ease certain geometrical constraints. For example, Ungerer et al. (2000) used vapour pressures, heats of vaporization, and liquid densities but assumed anisotropy of the pseudo-atoms. McKnight (2005) has used vapour pressures, liquid densities, vapour densities and heats of vaporization but allowed the equilibrium bond-lengths and bond-angles to be different from the experimentally measured values.

Vapour Pressures

In Figure 7.2 the Clausius-Clapeyron plots have been given for the n-alkanes. Not too surprising is the fact that both NERD and TraPPE-UA consistently over-predict the vapour

pressures. This is most likely due to the over-prediction of the vapour densities. From Equation 2.23 one can see that the ideal contribution to the vapour pressure is $\rho_{vap}k_BT$. Thus, if the vapour density is over-predicted then it is very likely that the vapour pressure will be over predicted as well. The NERD force field predictions appear to be slightly better than those of TraPPE-UA for the n-octane. This possibly due to the higher LJ-well depth (ϵ/k_B) but as the chain-length increases its effect is overshadowed by the LJ-well depth of the CH_2 pseudo-atom. Since the LJ-well depths for the CH_2 pseudo-atom are effectively the same for both NERD and TraPPE-UA, as the chain-length increases so the vapour pressures become very similar. It appears that as the vapour pressure predictions improve slightly as the temperatures approach critical. This can be explained by considering the system sizes involved. For the simulations at these higher temperatures more molecules would be in the vapour box due to the much higher vapour density. The higher number of molecules leads to a better prediction of the vapour pressure. Unfortunately, it is not possible to simulate vapour boxes of this size for the lower temperatures because the number of molecules that would be required in the liquid boxes would be far to high to allow practical simulations.

Critical Properties

The estimated normal boiling points for the n-alkanes were under-predicted. This trend follows directly from the fact that the vapour pressures were consistently over-predicted over the whole temperature range. However, even with this under-prediction the normal boiling points are still predicted to within about 3% of experimental values. The critical temperature estimations were to within 2% of experimental values. Figure 8.7 shows the ratio of the estimated normal boiling points and critical temperatures to their respective experimental values. From Figure 8.7 one can see that the deviation of the estimated normal boiling points is fairly consistent between 3% and 1% and this is likely to remain the case for longer n-alkanes. While the estimated normal boiling points were fairly consistent, there was little consistency with regard to the over- or under-prediction of the critical temperatures. They ranged from an under-prediction of about 0.5% up to an over-prediction of about 2%. The most likely cause is the accuracy to which the simulated densities could be calculated to. As the temperature neared critical so the standard deviations increased and confidence in the density results decreased. The simple regression method used was also shown to introduce a 1–2% error in results. For these reasons

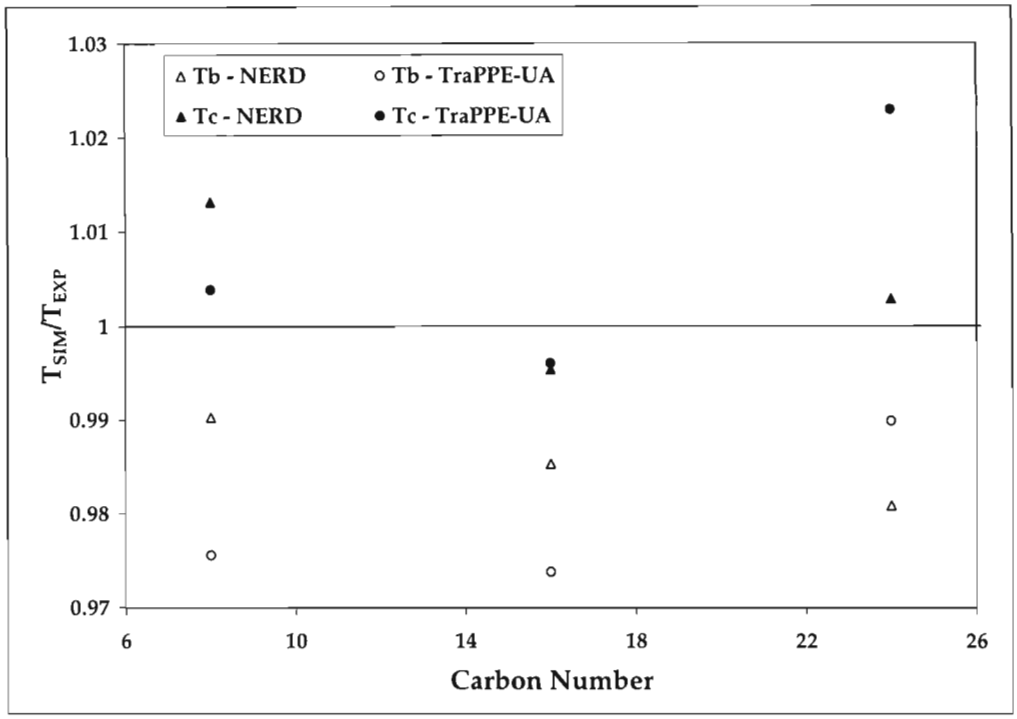

Figure 8.7: Ratio of simulated to experimental critical temperatures (filled symbols) and normal boiling points (open symbols) versus carbon number for selected n-alkanes using NERD (triangles) and TraPPE-UA (circles).

one cannot expect the predicted critical temperatures to be accurate to better than 1 or 2%. Martin and Siepmann (1998) argued that the trends presented here for both the normal boiling points and critical temperatures was probably not due to a particular set of force field parameters.

In Siepmann et al. (1993) the authors were able to show that the critical density increased to a maximum around ethane and then began to decrease. Figure 8.8 show the chain-length dependence of the critical density for the NERD and TraPPE-UA force fields with experimental data. One can see that NERD predicts the critical density very well but TraPPE-UA slightly over-predicts the critical densities. The over-prediction by TraPPE-UA of the critical densities is most likely due to the fact that the vapour-liquid coexistence curve is slightly shifted to the left (i.e both the vapour and liquid densities are over-predicted).

8.5.2 1-Alkenes

The second hydrocarbon group investigated were the 1-alkenes. This group was chosen because of its similarity to the n-alkanes. Both TraPPE-UA (Wick et al., 2000) and NERD

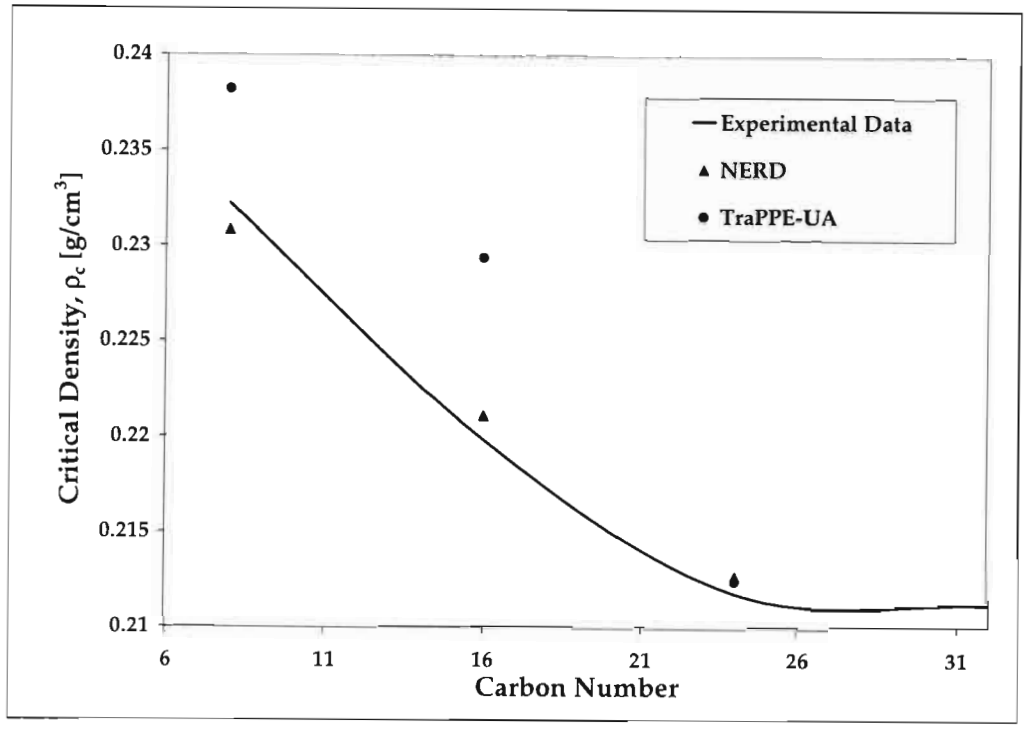

Figure 8.8: Critical densities versus the chain length for the simulated n-alkanes.

(Nath et al., 2001*a*) parameterized the the required pseudo-atoms during the early part of this decade. While some work has been published for the shorter 1-alkenes, it is important to note that there have been no publications for 1-alkenes longer than C_8. The simulation results for three of the 1-alkenes simulated during this study have been presented, viz. 1-octene, 1-decene and 1-hexadecene.

Vapour-Liquid Coexistence Curves

In Figure 7.3 the vapour-liquid coexistence curves for the simulated 1-alkenes have been shown. The lack of experimental or published simulation data makes the analysis of these data difficult. At the lower temperatures for the 1-octene simulations it appears that the curves follow the same trends as the n-alkanes - NERD under-predicts the liquid density and TraPPE-UA over-predicts the liquid density. Again, the likely cause of this is also the difference in the LJ-sizes of the pseudo-atoms. At temperatures closer to critical temperature the liquid coexistence curves predicted are shallow - the liquid densities are under predicted for the temperatures close to T_c. Interestingly, the Scaling/Rectilinear curves, which were fitted to the available experimental vapour-liquid coexistence data, followed the trends shown by simulations. Thus, it is possible that the experimental

data is in error. However, without further experimental investigations the exact cause of the discrepancy is uncertain. For the longer 1-alkenes, 1-decene and 1-hexadecene, the NERD predicts a lower liquid density compared with TraPPE-UA. While there was little density data to compare these coexistence curves to it is likely that the curves are fairly accurate. From these curves one can see that the different, non-polar functional group (CH_2=CH-) does not affect the vapour-liquid curves substantially for the longer 1-alkenes. The same trends for NERD and TraPPE-UA have been shown to occur for the n-alkanes and 1-alkenes and this shows the dominance of the CH_2 pseudo-atom on the long-chain data. It may appear that the Scaling/Rectilinear curves, using $\gamma = 0.36$, are unable to accurately describe the temperature region near to critical, particularly if one considers the "flattened" regions of the curves for 1-decene between 0.05–0.1 g/ml (vapour) and 0.4–0.5 g/ml (liquid). However, from the accuracy of the critical properties and the appropriate shape of the curves intersecting with the critical points it is likely due to a deficiency in the plotting of the curves.

Vapour Pressures

The Clausius-Clapeyron plots for the 1-alkenes have been shown in Figure 7.4. There was experimental vapour pressure data available for the 1-alkenes up to a pressure of 1atm. As with the n-alkanes, the vapour pressures are consistently over-predicted. However, is appears that the curves are slightly closer to the experimental curves than in the case of the n-alkanes. Since there was effectively no vapour density data available it is impossible to test how much of the over-prediction is due to the vapour density and how much due to the pseudo-atom parameters. If one considers the Scaling/Rectilinear curves it appears that the vapour densities are more closely estimated than they were for the n-alkanes. If this is case, then the over-prediction of the vapour pressures for the 1-alkenes is largely due to the intermolecular interactions between the CH_2 pseudo-atoms. As with the n-alkanes the vapour pressure curves improved in accuracy as the temperature increased to critical. Again, this is largely due to the more molecules that could be simulated in the vapour boxes.

Critical Properties

From the discussion in Section 8.3 and the trends shown in this section, one would expect the critical properties to follow the same trends as the n-alkanes. In fact, from Figure 8.6

one would expect the critical properties to be very similar to those of the n-alkanes. In Figure 8.9 the ratio of the estimated critical temperature and normal boiling points are shown. As expected, the normal boiling points are consistently under-predicted by both NERD and TraPPE-UA. Again, this follows directly from the fact that the vapour pressures are consistently over-predicted throughout the whole temperature range investigated. From Figure 8.9 it appears that there is no chain-length dependence for the normal boiling points. Neither NERD or TraPPE-UA consistently predict the normal boiling points more accurately, thus, the cause of the 1–3% difference between predicted and experimental data is likely due to the estimation procedure. The estimated criti-

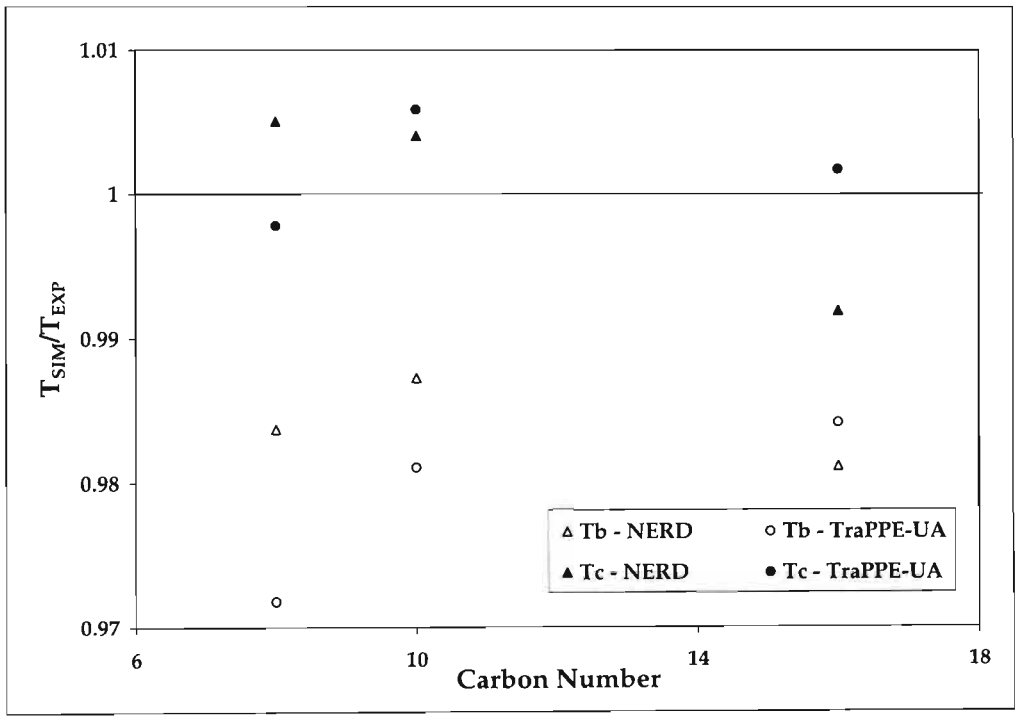

Figure 8.9: Ratio of simulated to experimental critical temperatures (filled symbols) and normal boiling points (open symbols) versus carbon number for selected 1-alkenes using NERD (triangles) and TraPPE-UA (circles).

cal temperatures were accurately predicted to within 1% of experimental data. The fact that the critical temperatures have been well estimated is another reason why the shape of the coexistence curves are probably correct. The critical densities for 1-octene and 1-decene are also fairly accurately estimated which again leads to the conclusion that the Scaling/Rectilinear curves are appropriate. The critical density for 1-hexadecene, which was obtained from the KDB, was substantially higher than the estimated values. The most reasonable possibility is that the value from the KDB is too high. There are a

number of reasons for concluding this. Firstly, if one considers critical densities for the n-alkanes, one can see that as the chain-length increases so the critical density decreases. This same sort of trend is followed by the simulation results but not for the experimental data if one considers 1-hexadecene. Secondly, using the available DDB data (up to 1-decene), the estimated critical density for 1-hexadecene should be about 0.238 g/ml (3% lower than the KDB value). The data is given the KDB without references so it is uncertain how exactly the value they give was obtained. However, it likely that the value given in the KDB is too high.

8.5.3 1-Alcohols

Many of the simulations performed for the 1-alcohols were novel. As with the 1-alkenes there appears to be no publications of simulations involving 1-alcohols longer than C_8. While there was more experimental data available for the longer 1-alcohols compared with the 1-alkenes, in many cases the data did not extend past about 550–600K. No experimental or simulation vapour-liquid coexistence data could be found for 1-eicosanol. The KDB did have some of the critical properties and parameters for the Antoine equation but the source of these data is uncertain. The reason for the lack of data is that for these longer 1-alcohols there are difficulties involved with experimentally measuring these data due to thermal degradation. However, as discussed earlier in this chapter, this data is still required. Since complex iterative equations of state type methods are typically required for alcohols the option of using MC simulations becomes attractive. Unfortunately, due to time limitations these more accurate methods could not be investigated during this study. Only two force fields were used for the n-alkane and 1-alkene simulations. A third one, OPLS-UA which pre-dates NERD and TraPPE-UA by over a decade, was also used for the simulation of the long-chain 1-alcohols.

Vapour-Liquid Coexistence Curves

The vapour-liquid coexistence curves have been shown separately in Figures 7.5 (C_8), 7.7 (C_{12}), 7.9 (C_{16}) and 7.11 (C_{20}). They have been shown separately because the simulation results were so different for the experimental data and the NERD and TraPPE-UA predictions. There are a number of interesting and important aspects when considering these figures. The most obvious one is the fact that OPLS-UA has performed so badly for all the 1-alcohols simulated. It over-predicts the liquid densities and under-predicts

the vapour densities substantially. This is most likely caused by the OPLS-UA LJ-well depth being 30% larger than the NERD or TraPPE-UA LJ-well depths. This causes the attraction at closer intermolecular regions to be greater. One would expect, that if this is the case, for the effect to become more noticeable as the chain length increases since the longer molecules have more of a tendency to attract other molecules. As the chain-length increases so the over-prediction of the liquid densities also increases, thus it is likely that the large LJ-well depth is the major cause. This attraction does not affect the vapour phase to any great extent because the molecules are spread further apart. The fixed bond-angle also removes the flexibility which is probably required to correctly simulate both the liquid and vapour phases correctly. What is more, the OPLS-UA force field was parameterized for predicting the liquid density of 1-alcohols (shorter than C_8) at 25°C. At temperatures so much lower than those used in this study it is difficult to predict the errors that would propagate through to these higher temperatures. It seems however that the parameterization of the CH_2 pseudo-atom is the major cause of the incorrect predictions of OPLS-UA. The LJ-sizes do not contribute to the over-prediction in this case because the LJ-sizes for all the relevant pseudo-atoms are very similar. Secondly, one can see that while both NERD and TraPPE-UA perform much better than OPLS-UA, they both seem to generate coexistence curves that are too narrow (i.e. the vapour densities are over-predicted and the liquid densities are under-predicted) and too shallow (i.e. the critical temperatures are under-predicted). As the chain-length increases so the shallowness of the coexistence curves appears increase. It has been suggested by Chen et al. (2001) that these could be due to the functional forms used rather than any particular parameter set. This possibly could be the case since both NERD and TraPPE-UA show the same trend and both are slightly different in their parameterizations. Interestingly TraPPE-UA consistently predicts a lower liquid density and a higher vapour density which is the opposite trend to the one observed for the n-alkanes and 1-alkenes. This is most likely due the OH functional group, which has been shown (Figure 8.6) to affect the properties substantially. The LJ-size for the O pseudo-atom for TraPPE-UA is about 1% larger than that of NERD. Thus, even though NERD does not use a ghost H pseudo-atom whereas TraPPE-UA does, the TraPPE-UA should slightly under-predict the liquid density compared with NERD. This slight under-prediction compared causes a slight over-prediction for the vapour density.

Vapour Pressures

In Figures 7.6, 7.8, 7.10 and 7.12 the Clausius-Clapeyron plots have been shown for the 1-alcohols simulated during this study. As with the vapour-liquid coexistence curves the vapour pressure plots have been shown separately due to the large variation in the predicted vapour pressures. Many of the trends noticed for the vapour-liquid coexistence curves directly affect the vapour pressure trends. The most obvious trend is the under-prediction of the vapour pressure by the OPLS-UA force field. This is due to the fact that the vapour densities have been substantially under-predicted. This vapour pressure under-prediction increases as the chain-length increases. This follows from the fact that the vapour densities are increasingly under-predicted as the chain-length increases. There does not appear to be any dependence for the OPLS-UA vapour pressures on the temperature of the simulation. This most likely due to the fact that OPLS-UA so substantially under-predicts the vapour densities and pressures that even near the critical temperature there well still not many molecules in the vapour box. Thus, the the slight variations noticed in NERD and TraPPE-UA for temperatures near critical may still occur for OPLS-UA only at much higher temperatures. Both NERD and TraPPE-UA over-predict the vapour pressure as usual. This is expected since both force fields over-predict the vapour densities. Since TraPPE-UA consistently predicts slightly higher vapour densities compared with NERD it follows that the vapour pressures predicted using TraPPE-UA are slightly higher than those predicted using NERD. AS the chain-length increased so the predictions of the KDB Antoine equation begin to over-predict the vapour pressures. This fact will be considered when discussing 1-eicosanol.

Critical Properties

It has already been stated that the both NERD and TraPPE-UA vapour-liquid coexistence curves were too shallow. Both the normal boiling points and the critical temperatures were consistently under-predicted by NERD and TraPPE-UA. The OPLS-UA force field however, consistently over-predicted both the normal boiling points and critical temperatures. These trends can be clearly seen in Figure 8.10 where the ratio of the estimated normal boiling points and critical properties to the experimental values has been shown. From Figure 8.10 one can see that OPLS-UA consistently over-predicts the normal boiling points and critical temperatures by at least 10%. There is a clear chain-length

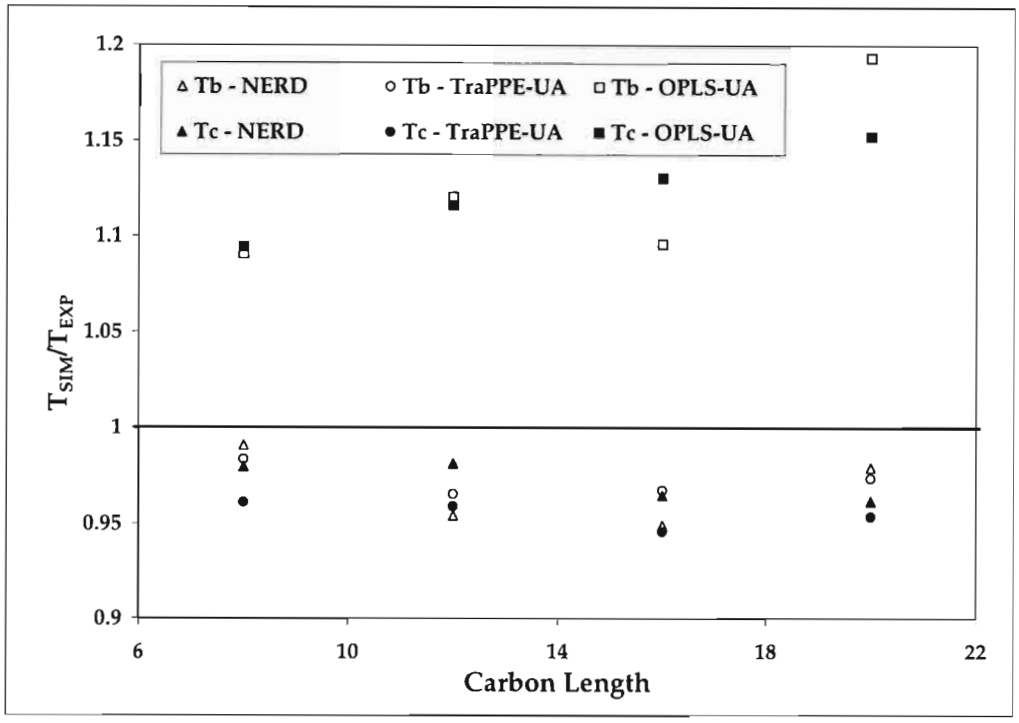

Figure 8.10: Ratio of simulated to experimental critical temperatures (filled symbols) and normal boiling points (open symbols) versus carbon number for selected 1-alkenes using NERD (triangles), OPLS-UA (squares) and TraPPE-UA (circles).

dependence which implies that the CH_2 pseudo-atom has been badly parameterized for the simulation of long-chain alcohols. It is possible that fixed bond-angle nature of the OPLS-UA molecules affects these properties as well. Unfortunately, without further investigation into the particular parameters it is not possible to quantify the effects due to any particular aspect of the force field. While both NERD and TraPPE-UA perform better than OPLS-UA, they still under-predict the normal boiling points and critical temperatures by as much as 5% for the longer 1-alcohols. There does not appear to be any chain-length dependence on these properties for NERD or TraPPE-UA. However, the general quality of the predictions decreases as the chain-length increases. This is likely due to the uncertainties introduced by simulating such long polar molecules and the uncertainties in the experimental data. The NERD force field seems to generally predict the 1-alcohol critical properties slightly better than the TraPPE-UA force field. This is probably due to a number of aspects within the force field parameterization, but one interesting possibility is the fact that NERD does not use a ghost pseudo-atom for the hydroxyl H. While TraPPE-UA predicts the thermophysical properties adequately, it is possibility that a more realistic pseudo-atom for the hydroxyl H is required to fine-tune

the predictions.

1-Eicosanol

The simulations of 1-eicosanol presented the most difficulties during this study. These simulations took the longest (even longer than the n-tetracosane simulations due to the Coulomb interactions which were considered) and there was effectively no experimental data to guide the choice of simulation conditions. The "experimental" critical properties presented in this dissertation were obtained from KDB and appear to extrapolated from shorter 1-alcohol data. Never the less, there is high confidence in the results obtained in this study for 1-eicosanol. The coexistence curves predicted by OPLS-UA follow the same trends of over-predicting the liquid density and under-predicting the vapour density and pressure. The normal boiling point and critical temperature predicted are the worst yet for OPLS-UA, being over-predicted by 15–20%. However, the NERD and TraPPE-UA force fields perform fairly well. The vapour-liquid coexistence curves are probably too narrow as per usual, but the estimated critical densities are still estimated to within reasonable limits (about 10% for NERD and 1% for TraPPE-UA). The KDB Antoine equation is probably over-predicting the vapour pressure as this trend has been shown from the vapour pressures of 1-dodecanol and 1-hexadecanol. Thus, it is likely that the vapour pressures are more over-predicted than they appear in Figure 7.12.

Hydrogen Bonding

The last aspect of the 1-alcohols that will be briefly discussed is hydrogen bonding. In Figure 8.11 the hydrogen and oxygen pseudo-atoms are shown for the liquid phase of 1-octanol at 550K. From this snapshot it is clear that some aggregation is occurring and the likely cause of this is hydrogen bonding. From this single aspect of the simulation box (with the boundaries shown) it is difficult to visually determine the average number of bonds forming. However, by using software such as RasMol (Bernstein, 2001), a free 'pdb'-imaging package[§], it is possible to rotate the box about any axis. Thus, it was possible to estimate that hydrogen bonding was occurring between typically two or three 1-octanol molecules. Further analysis of the output file `towhee_movie` would yield more information regarding the radial distribution functions (such as Figure 8.4) and oxygen–oxygen separations. This information would allow one to further investigate

[§]The RasMol software was also used to produce Figure 8.11.

Figure 8.11: Snapshot of the liquid phase of 1-octanol at 550K. Only the hydrogen and oxygen pseudo-atoms are shown, the carbons have been removed for clarity and the box boundaries have not been included.

the aggregation that occurs. Chen et al. (2001) has already investigated the hydrogen bonding of alcohols and further investigation into this aspect was not possible during this study due to time constraints.

8.5.4 Binary n-Alkane Mixtures

The major focus of the study has been the determination of pure component thermophysical properties. Due to the lack of experimental data for long-chain vapour-liquid equilibria and inherent difficulties associated with binary MC simulations (see Chapter 6.2.1) this aspect was only briefly investigated during this study. Two systems were simulated using the NERD force field, both are n-alkane mixtures. The NERD force field was used since it has been established previously, by McKnight et al. (2002), that NERD performed slightly better than TraPPE-UA for shorter hydrocarbon mixtures. The experimental data used in this section was obtained from KDB and has given in Appendix D.

P-x-y Curves

The pressure-composition curves for the two mixtures investigated during this study have been shown in Figures 7.13 (n-pentane + n-octane) and 7.15 (n-hexane + n-hexadecane). One can see that the P-x curves are generally over-predicted. The P-y curve in Figure 7.13 is also over-predicted slightly. For the n-hexane + n-hexadecane system, there was no experimental P-y data so this curve could not be compared. However, considering the n-pentane + n-octane system, it is likely that the P-y curve will be over-predicted. These results come directly from the fact that NERD over-predicts the vapour pressures. This was shown to always be the case for the pure component systems investigated and the trend is likely to continue for any non-polar mixtures investigated. Still, the curves predicted by NERD are fairly accurate. The mixing rules used to combine the parameters are the same as those used for the pure component simulations, Lorenz-Berthelot.

x-y Curves

Only the x-y curve for the n-pentane + n-octane system was given (Figure 7.14) since there was no experimental vapour composition data available for the n-hexane + n-hexadecane system and thus nothing to compare the simulation curve to. The x-y curve for n-pentane + n-octane was also predicted adequately. For the lower compositions the curve was slightly under-predicted and for the higher compositions it was slightly over-predicted. This effect is due to the dominance of the different vapour pressures on the system. Typically these relatively ideal curves are well represented no matter which force field is used since the errors in the vapour pressures tend to cancel out when comparing them.

Practicality of Binary Gibbs Ensemble Simulations

Clearly there is little point to the use of MC simulation in prediction of these simple mixtures. Simple predictive methods such as Raoult's Law or the modified Raoult's Law are quite adequate for these fairly ideal systems. More exotic systems have been investigated by many other researches. For example, Chen et al. (2000) simulated alkane-alcohol mixtures and Nath (2003) investigated mixtures of H_2S with alkanes. However, not without difficulties, and in many publications of binary simulations the authors have avoided the low concentrations and conditions where the phase envelops are thin. Under these

equilibrium conditions it is extremely difficult to simulate the systems using standard NVT-Gibbs techniques. Clearly, other more versatile methods of simulating mixtures are required. Currently the Histogram Reweighting techniques appear to be a very good possibility (see Chapter 6.2.2). Escobedo (1998) has presented other pseudo-ensembles for predicting multi-component phase equilibria. More recently, McKnight (2005) has introduced alternative simulation ensembles for multi-component vapour-liquid equilibria which avoid or solve many of the difficulties associated with the NPT-Gibbs ensemble and the pseudo-ensembles of Escobedo (1998).

Chapter 9

Conclusions

The aim of this study was to investigate the simulation of long-chain linear hydrocarbons using standard Monte Carlo (MC) methods. These simulations were performed over a range of temperatures from below the normal boiling points to just below the critical temperatures. These vapour-liquid coexistence data were then used to estimate the normal boiling points and the critical properties. These simulations were done using three transferable force fields, viz. NERD, OPLS-UA and TraPPE-UA. The results of these simulations were then compared with available experimental data.

9.1 Literature Data

It was found there is was very little thermophysical experimental data for heavy hydrocarbon compounds. Thus, in absence of such data there is a great need for accurate predictive methods. The method investigated in this study, Monte Carlo simulation proved to be a possible method. The key advantages that MC simulation has over other methods,such as EOS methods, are that no critical properties are required and that compounds may be extensively investigated under conditions that currently cannot be investigated experimentally.

9.2 Potential Models

An important aspect of MC simulations that must be improved are the force fields. While both NERD and TraPPE-UA were shown to be clearly superior to OPLS-UA for the the prediction of the 1-alcohol systems there is still substantial improvements in accuracy

required before these force fields will compete with other predictive methods. It is possible that different functional forms will be required to more accurately represent the inter- and intramolecular interactions since it was shown that much of the inaccuracy was not due to any particular parameter set.

The similarity between NERD and TraPPE-UA predictions for these long-chain hydrocarbons was shown to be largely due the very similar CH_2 pseudo-atom LJ parameters. Thus, while it is vital that the other pseudo-atoms be accurately represented in order to predict short-chain vapour-liquid equilibria, for long-chains one must ensure that the CH_2 pseudo-atom is accurately represented. Due to the complexity associated with MC simulations it is not possible to quantify the accuracy of particular pseudo-atoms. However, the use of bond-stretching by NERD appears to be unnecessary since the results were not appreciably improved by its use, yet the time required for the simulation did increase due to added CBMC trials that were needed.

9.3 Simulation Results

Both NERD and TraPPE-UA consistently over-predicted the vapour densities and pressures by 10–20% of experimental values for all the hydrocarbon groups investigated. The liquid densities were typically predicted to within 1–3% of experimental values. The large difference in accuracy between the two sets of properties can be attributed to the fact that both force fields were parameterized using only experimental liquid densities and not taking vapour densities and pressures into account. Clearly, more experimental data must be used when fitting these models. In order to use more experimental data in the fitting process, more flexibility in geometries will be required.

Typically, the major effect of the chain-length was to increase time required to simulate the hydrocarbons to obtain results with reasonable standard deviations. The more modern force fields used during this study, NERD and TraPPE-UA, appear to be adequately parameterized for these hydrocarbons. The older force field, OPLS-UA, however showed a probable weakness in the parameterization of it's CH_2 pseudo-atom since the quality of the calculated thermophysical properties decreased with increasing chain-length.

9.4 Estimated Critical Properties

The estimated normal boiling points and critical properties were shown to be acceptable. The estimated normal boiling points were typically within 2–5% and the critical temperatures and densities were within 1–5% of experimental values. A simple method of estimating these properties was used during this study, and a more advanced method could have been employed in conjunction with an alternative ensemble (such as finite size analysis with the grand-canonical ensemble) in order to get better predictions. However, results based on these methods would still only be as accurate as the force fields allowed and thus not likely to be much better than those obtained in this study.

9.5 Industrial Viability

It has been shown that standard MC simulation techniques may be used to simulate long-chain hydrocarbons. However, the time required is between 30 and 120 hours for a single vapour-liquid coexistence point. Thus, more development is required on algorithm techniques and the speed of computers needs to increase substantially before MC simulation of long-chain hydrocarbons become an industrially viable method.

Chapter 10

Recommendations

The molecular simulation of large hydrocarbons is currently of great interest. Linear long-chain hydrocarbons are only a small aspect of this very large field. There were many other aspects of equal interest that were not investigated during this study due to time constraints. A brief list of some of these areas are:

- MC simulation of long-chain branched hydrocarbons,
- MC simulation of cyclic or aromatic hydrocarbons,
- Prediction of binary and ternary vapour-liquid equilibria for these large hydrocarbons,
- Solubility of industrially important gases (such as CO_2 or H_2S) in large hydrocarbons,
- Investigation into other simulation techniques for simulating these large hydrocarbons,
- Molecular dynamical simulations could be investigated for the prediction of dynamical properties such as viscosity.

Many of these topics are currently been studied or have already been partially investigated by other researchers. The first three areas listed above were slightly investigated during this study. They all may be considered natural progressions from the the work performed in this study. However, each of these areas is very large, one simply has to consider the number of additional compounds or systems that any one of these areas introduces. Of the the three, the investigations into branched chains seems the most

viable since no new techniques (other than coupled-decoupled CBMC) are particularly required. The other two, however, introduce a number of additional difficulties. The cyclic or aromatic hydrocarbons are notoriously difficult to grow and insert effectively and while research is currently being performed to solve this difficulty, there are still no simple solutions. The binary and ternary vapour-liquid equilibria introduces many of the difficulties discussed for when one uses the NPT-Gibbs ensemble. Also the sheer number of simulations required to generate a full phase envelope, no matter which MC method one uses, increases rapidly.

The last three areas listed above would be particularly interesting topics to investigate. The solubility of gases in hydrocarbons is industrially important. And through the use of MC simulations one could investigate the types of bonding that occur, such as aggregation, and vapour-liquid equilibria associated with these mixtures. However, the same difficulties arise when attempting to simulate these mixtures as with the hydrocarbon mentioned earlier. Clearly, more affective MC methods are required and research into this would greatly advance the field of MC simulation. The final area suggested comes via the fact that all the properties calculated using MC simulation methods are equilibrium properties. In many cases time-dependent properties such as viscosity or reaction rates may be of great importance.

Appendix A

Potential Model Parameters

This appendix lists the model parameters for the force fields used during this study. The parameter lists presented here are by no means exhaustive. Only the parameters specifically used in this study have been shown. Martin (2004*a*) contains an expansive list of force fields and their implementation. The values for these force field parameters may be found in the following references. For the NERD parameter values:

- n-alkanes (Nath, Escobedo and de Pablo, 1998),
- 1-alkenes (Nath et al., 2001*a*),
- 1-alcohols (Khare et al., 2004).

And, for the OPLS-UA parameter values:

- n-alkanes and 1-alkenes (Jorgensen et al., 1984),
- 1-alcohols (Jorgensen, 1986).

Finally, for the TraPPE-UA parameter values:

- n-alkanes (Martin and Siepmann, 1998),
- 1-alkenes (Wick et al., 2000),
- 1-alcohols (Chen et al., 2001).

Bond	r_0 [Å]	k_r/k_B [K]
CH_x-CH_y	1.54	168380
CH_x-OH	1.43	198448
O-H	0.961	312706
CH_x=CH_y	1.34	48250

Table A.1: NERD bond-stretching parameters.

Bend	θ_0 [deg.]	k_θ/k_B [K]
CH_x-(CH_2)-CH_y	114	45703
CH_x-(CH_y)-O	108	60136
CH_x-(O)-H	107.5	27662
CH_x=(CH)-CH_y	124	31250

Table A.2: NERD bond-bending parameters.

Torsion	c_0/k_B [K]	c_1/k_B [K]	c_2/k_B [K]	c_3/k_B [K]
CH_x-(CH_2)-(CH_2)-CH_y	0.0	355.04	-68.19	791.32
CH_x-(CH_2)-(CH_2)-OH	232.00	356.37	-472.19	1119.9
CH_x-(CH_2)-(O)-H	0.0	359.25	59.053	220.82
CH_x=(CH)-(CH_2)-CH_y	47.97	86.31	-109.71	282.08

Table A.3: NERD torsional parameters.

United-Atom	ϵ/k_B [K]	σ [Å]	Charge, q
CH_3	104	3.91	-
CH_2 (sp^3)	45.8	3.93	-
CH_2 (sp^2)	92.5	3.72	-
CH (sp^2)	46	3.77	-
α-CH_2 (sp^3)	45.8	3.93	+0.290
O	108	2.98	-0.710
H	3.89	0.98	+0.420

Table A.4: NERD non-bonded parameters.

Bond	r_0 [Å]
CH_x-CH_y	1.53
CH_x-OH	1.43
O-H	0.945

Table A.5: OPLS-UA bond-stretching parameters.

Bend	θ_0 [deg.]
CH_x-(CH_2)-CH_y	112.0
CH_x-(CH_y)-O	108.0
CH_x-(O)-H	108.5

Table A.6: OPLS-UA bond-bending parameters.

Torsion	c_0/k_B [K]	c_1/k_B [K]	c_2/k_B [K]	c_3/k_B [K]
CH_x-(CH_2)-(CH_2)-CH_y	0.0	355.0	-68.18	791.3
CH_x-(CH_2)-(CH_2)-OH	0.0	176.6	-53.34	769.9
CH_x-(CH_2)-(O)-H	0.0	209.8	-29.18	187.9

Table A.7: OPLS-UA torsional parameters.

United-Atom	ϵ/k_B [K]	σ [Å]	Charge, q
CH_3	88.06	3.905	-
CH_2 (sp^3)	59.38	3.905	-
α-CH_2 (sp^3)	59.38	3.905	+0.265
O	85.5	3.07	-0.700
H	0.0	0.000	+0.435

Table A.8: OPLS-UA non-bonded parameters.

Bond	r_0 [Å]
CH_x-CH_y	1.54
CH_x-OH	1.43
O-H	0.945
CH_x=CH_y	1.33

Table A.9: TraPPE-UA bond-stretching parameters.

Bend	θ_0 [deg.]	k_θ/k_B [K]
CH_x-(CH_2)-CH_y	114	62500
CH_x-(CH_y)-O	109.47	50400
CH_x-(O)-H	108.5	55400
CH_x=(CH)-CH_y	119.7	70420

Table A.10: TraPPE-UA bond-bending parameters.

Torsion	c_0/k_B [K]	c_1/k_B [K]	c_2/k_B [K]	c_3/k_B [K]
CH_x-(CH_2)-(CH_2)-CH_y	0.0	335.03	-68.19	791.32
CH_x-(CH_2)-(CH_2)-OH	0.0	176.62	-53.34	769.93
CH_x-(CH_2)-(O)-H	0.0	209.82	-29.17	187.93
CH_x=(CH)-(CH_2)-CH_y	688.5	86.36	-109.77	-282.24

Table A.11: TraPPE-UA torsional parameters.

United-Atom	ϵ/k_B [K]	σ [Å]	Charge, q
CH_3	98	3.75	-
CH_2 (sp^3)	46	3.75	-
CH_2 (sp^2)	85	3.95	-
CH (sp^2)	47	3.73	-
α-CH_2 (sp^3)	46	3.75	+0.265
O	93	3.02	-0.700
H	0	0.000	+0.435

Table A.12: TraPPE-UA non-bonded parameters.

Appendix B

Enabling Technologies

B.1 Yoda: The Beowulf Cluster

B.1.1 Hardware

Yoda consists of 1 master-node and 19 slave-nodes. The basic hardware specs on each node are as follows:

- Athlon AMD 1.2GHz
- 256MB RAM
- 100Mbit network card
- 20GB Hard drive (80GB for master node)

The master-node also has a monitor and CD-ROM drive connected for installation and debugging requirements. The Yoda system is connected to the University of KwaZulu-Natal LAN through a second 100Mbit network card in the master-node. All the slave-nodes are connected to the master node using a 100MB/s switching hub. Industrially crimped network cables were used for node-to-hub cabling in order to reduce the network-latency within Yoda. Figure B.1 shows a photograph of Yoda.

B.1.2 Software

Yoda is a Linux based Beowulf cluster. Currently it uses the Rocks 3.1.0 (Matterhorn) distribution. The system has been configured to act as a master-node connected to a

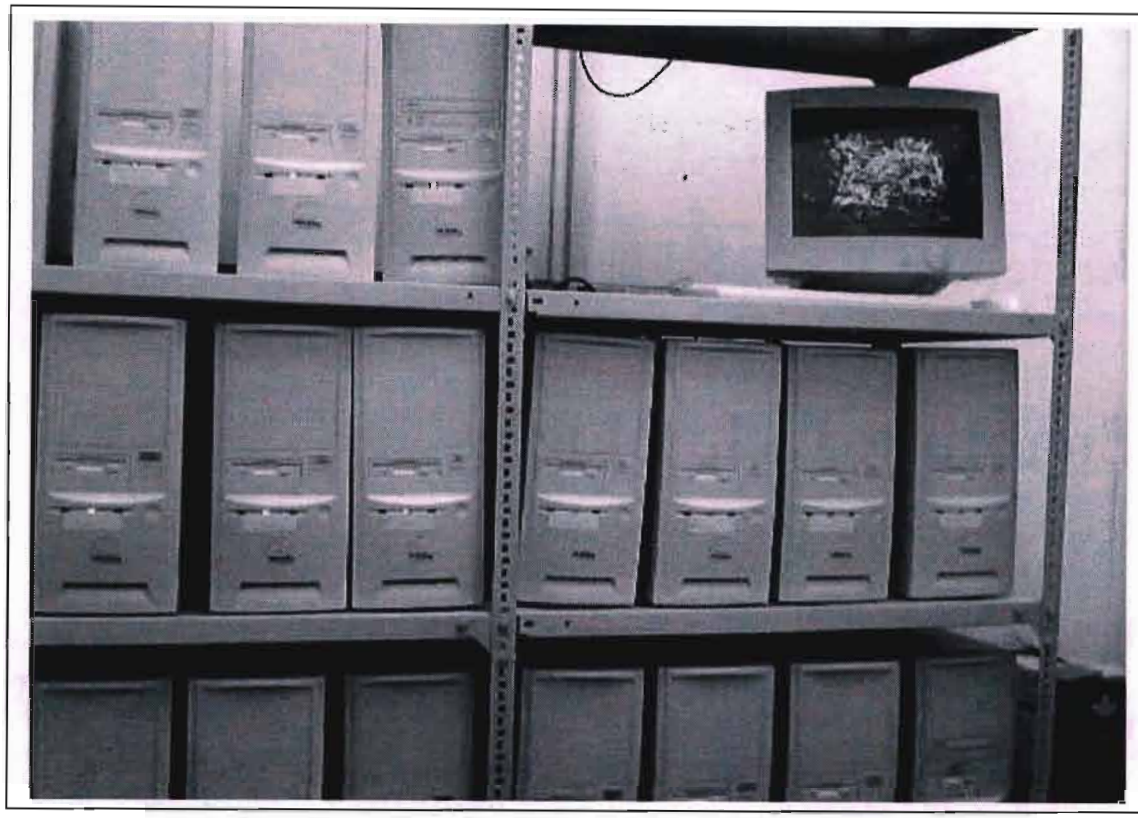

Figure B.1: Yoda, the Beowulf Cluster.

number of slave-nodes which perform the assigned computational operations independently to one-another. A job submission daemon called Grid-engine was used to submit jobs on the master-node. These jobs were then distributed by Grid-engine to an available slave-node.

A number of BASH scripts were used during this study to speed up the processing of the huge amounts of data produced during the simulations. Appendix E shows the BASH script that was used to extract the run-time pressures, densities, energies, etc.

Early investigations by McKnight (2003) into the viability of parallel processing of Monte Carlo simulations was found to be largely dependent on the system size. And, since the systems dealt with in this study consist of a relatively small number of molecules ($N < 500$), parallel processing techniques was not used.

B.2 Monte Carlo Simulation Codes

Two semi-related molecular simulation codes were used to simulate the hydrocarbon systems during this project. They are BIGMAC and TOWHEE . Both codes are loosely

based on work originating from Siepmann et al. (1993). The BIGMAC code was used for the non-polar simulations (i.e. the n-alkanes, 1-alkenes and binary n-alkanes) and TOWHEE code was used for the polar simulations (i.e. the 1-alcohols). Both BIGMAC and TOWHEE are coded in Fortran-77. An excellent overview of Fortran-77 is that of Page (1988).

B.2.1 BIGMAC

The serial version BIGMAC code was originally developed by T.J.H. Vlugt. It was developed to investigate the adsorption of alkanes into zeolites but was able to simulate systems without zeolites using the Gibbs ensemble. It was subsequently updated by T. McKnight to be more applicable to the simulation of vapour-liquid coexistence. A public version of BIGMAC may be found at `http://molsim.chem.uva.nl/bigmac/`.

Available Ensembles and Force Fields

The version of BIGMAC used during this study was able to simulate systems in the following ensembles:

- canonical (NVT) ensemble,
- isothermal-isobaric (NPT) ensemble,
- Gibbs ensemble (both NVT and NPT),
- grand-canonical (μVT) ensemble,

While it was possible to add new force fields with the same functional form (Lennard-Jones 12-6), only two force fields were present and used:

- NERD
- TraPPE-UA

B.2.2 MCCCS TOWHEE

The TOWHEE code was largely written, and now maintained by Marcus G. Martin. It is developed in collaboration with the Siepmann research group (University of Minnesota). Further information about TOWHEE and the links to downloading the software may be found at `http://towhee.sourceforge.net/`.

Available Ensembles and Force Fields

Using the most recent version of TOWHEE (version 4.4.2), one is able to perform simulations in the following ensembles:

- canonical (NVT) ensemble,
- isothermal-isobaric (NPT) ensemble,
- Gibbs ensemble (both NVT and NPT),
- grand-canonical (μVT) ensemble,

Since TOWHEE offers a very large number of force fields, the list given here will consist solely of those using the Lennard-Jones 12-6 functional form:

- Amber param96
- Aqvist ions
- Charmm22
- Charmm27
- ClayFF
- DACNIS United Atom
- DREIDING Dubbeldam *et al.* (alkanes and zeolites)
- Gromos 43A1
- Lastoskie *et al.* . N_2
- Lennard-Jones beads
- Lybrand-Ghosh-McCammon ions
- NERD
- OPLS-AA
- OPLS-UA
- SKS n-alkanes

- SMMK
- SPC-E water
- Sum *et al.* 2003
- TIP3P, TIP4P, and TIP5P water
- TraPPE-UA

Appendix C

Numerical Results

In this appendix the numerical coexistence densities and saturated vapour pressures are presented for the simulations performed during this study. Each table is for a particular hydrocarbon, with the results for every force field used. In all the tables the subscripts represent the statistical accuracy of the final digit(s).

C.1 n-Alkanes

In this section the numerical simulation results for n-octane, n-hexadecane and n-tetracosane are given. Only the NERD and TraPPE-UA force fields were used to simulate these hydrocarbons.

Force Field	T [K]	P [kPa]	ρ_g [g/ml]	ρ_l [g/ml]
NERD	340	17_4	$0.0007_?$	$0.6551_?$
	400	123_7	$0.0045_?$	$0.6028_?$
	460	468_{30}	$0.0160_?$	$0.5397_?$
	520	1285_{147}	$0.0484_?$	$0.4603_?$
TraPPE-UA	340	21_1	0.0009_1	0.669_1
	400	143_9	0.0052_3	0.615_2
	460	522_{14}	0.0187_4	0.551_2
	520	1423_{189}	0.054_5	0.464_8

Table C.1: Vapour-liquid coexistence simulation results for n-octane.

Force Field	T [K]	P [kPa]	ρ_g [g/ml]	ρ_l [g/ml]
NERD	450	5_1	0.0003_1	0.660_8
	500	23_9	0.0013_5	0.614_5
	550	129_{35}	0.585_5	0.006_1
	600	282_{47}	0.526_8	0.015_2
	650	618_{80}	0.460_9	0.035_4
TraPPE-UA	450	7_1	0.00045_6	0.671_1
	500	37_3	0.0021_1	0.632_2
	550	116_{10}	0.0062_5	0.589_3
	600	297_{27}	0.016_1	0.535_8
	650	660_{64}	0.039_4	0.47_1

Table C.2: Vapour-liquid coexistence simulation results for n-hexadecane.

Force Field	T [K]	P [kPa]	ρ_g [g/ml]	ρ_l [g/ml]
NERD	600	33_{12}	0.0023_9	0.58_1
	650	92_{10}	0.006_8	0.537_6
	700	239_{28}	0.016_2	0.48_1
	750	624_{188}	0.04_1	0.42_2
TraPPE-UA	600	30_7	0.0021_5	0.595_9
	650	87_{15}	0.005_1	0.549_9
	700	217_{38}	0.014_2	0.503_9
	750	470_{135}	0.035_8	0.44_1

Table C.3: Vapour-liquid coexistence simulation results for n-tetracosane.

C.2 1-Alkenes

Here the numerical simulation results for 1-octene, 1-decene and n-hexadecene are presented. Again, only NERD and TraPPE-UA force fields were used to simulate these hydrocarbons.

Force Field	T [K]	P [kPa]	ρ_g [g/ml]	ρ_l [g/ml]
NERD	390	105_{10}	0.0038_4	0.625_2
	400	138_{11}	0.0049_3	0.615_9
	450	434_{12}	0.0150_5	0.563_2
	480	747_{30}	0.0262_2	0.525_5
	510	1214_{82}	0.0445_8	0.481_2
	540	1871_{64}	0.075_4	0.422_3
TraPPE-UA	390	117_5	0.0042_1	0.633_2
	400	161_6	0.0057_2	0.624_1
	450	471_{15}	0.0166_7	0.569_2
	480	832_{44}	0.029_1	0.532_2
	510	1308_{42}	0.048_3	0.484_5
	540	2100_{171}	0.08_1	0.42_1

Table C.4: Vapour-liquid coexistence simulation results for 1-octene.

Force Field	T [K]	P [kPa]	ρ_g [g/ml]	ρ_l [g/ml]
NERD	350	5_1	0.00027_8	0.695_3
	400	37_3	0.0016_1	0.652_2
	450	146_{11}	0.0058_4	0.606_1
	500	407_{17}	0.0161_2	0.553_2
	550	929_{80}	0.038_4	0.486_9
TraPPE-UA	350	6_1	0.00029_7	0.702_2
	400	40_4	0.0017_2	0.660_2
	450	153_{17}	0.0062_7	0.614_2
	500	433_{24}	0.017_1	0.561_7
	550	1013_{52}	0.042_2	0.496_7

Table C.5: Vapour-liquid coexistence simulation results for 1-decene.

Force Field	T [K]	P [kPa]	ρ_g [g/ml]	ρ_l [g/ml]
NERD	500	32_6	0.0017_3	0.631_7
	550	107_{23}	0.005_1	0.58_1
	600	300_{50}	0.016_3	0.53_2
	620	398_{65}	0.021_5	0.50_2
	650	636_{101}	0.033_6	0.46_1
	670	862_{144}	0.050_9	0.40_8
TraPPE-UA	500	31_5	0.0017_2	0.638_6
	550	110_{22}	0.005_1	0.595_6
	600	280_{22}	0.015_3	0.543_9
	620	343_{22}	0.017_2	0.506_7
	650	582_{50}	0.030_3	0.47_1

Table C.6: Vapour-liquid coexistence simulation results for 1-hexadecene.

C.3 1-Alcohols

This is the final group of hydrocarbons for which pure component simulation results have be given. For 1-octanol, 1-dodecanol, 1-hexadecanol and 1-eicosanol the NERD, OPLS-UA and TraPPE-UA force fields were used.

Force Field	T [K]	P [kPa]	ρ_g [g/ml]	ρ_l [g/ml]
NERD	450	68_{12}	0.0023_4	0.699_6
	500	247_{58}	0.008_2	0.64_1
	550	766_{111}	0.026_3	0.58_1
	570	988_{161}	0.035_6	0.55_1
	590	1313_{210}	0.047_9	0.52_2
TraPPE-UA	450	77_9	0.0027_4	0.697_6
	500	265_{23}	0.0090_9	0.635_5
	550	815_{40}	0.029_2	0.579_6
	570	1151_{156}	0.042_7	0.54_2
	590	1730_{126}	0.066_9	0.506_4
OPLS-UA	450	15_5	0.0005_2	0.757_3
	500	80_{10}	0.0025_3	0.714_4
	550	303_{34}	0.009_1	0.668_6
	570	442_{35}	0.013_1	0.649_4
	590	560_{84}	0.017_2	0.625_5

Table C.7: Vapour-liquid coexistence simulation results for 1-octanol.

Force Field	T [K]	P [kPa]	ρ_g [g/ml]	ρ_l [g/ml]
NERD	500	60_{2}	0.0028_{1}	0.66_{1}
	530	197_{39}	0.009_{3}	0.62_{9}
	560	301_{85}	0.013_{3}	0.627_{9}
	590	456_{104}	0.021_{8}	0.57_{1}
	620	662_{250}	0.031_{9}	0.53_{1}
	650	925_{281}	0.04_{1}	0.49_{2}
TraPPE-UA	500	65_{17}	0.0030_{8}	0.670_{6}
	530	139_{28}	0.006_{1}	0.639_{7}
	560	272_{29}	0.012_{1}	0.610_{5}
	590	423_{36}	0.018_{2}	0.568_{8}
	620	714_{73}	0.034_{5}	0.535_{9}
	650	1062_{94}	0.054_{8}	0.47_{1}
OPLS-UA	500	13_{5}	0.0006_{2}	0.755_{5}
	530	22_{9}	0.0009_{3}	0.733_{4}
	560	48_{15}	0.0019_{6}	0.706_{5}
	590	72_{25}	0.002_{1}	0.684_{7}
	620	152_{44}	0.005_{1}	0.659_{6}
	650	253_{83}	0.010_{4}	0.627_{7}

Table C.8: Vapour-liquid coexistence simulation results for 1-dodecanol.

Force Field	T [K]	P [kPa]	ρ_g [g/ml]	ρ_l [g/ml]
NERD	550	64_{15}	0.0036_8	0.649_9
	600	221_{21}	0.012_1	0.601_7
	660	524_{62}	0.031_4	0.52_1
	680	672_{53}	0.039_3	0.50_2
	700	935_{142}	0.05_1	0.47_1
TraPPE-UA	550	52_4	0.0029_3	0.648_3
	600	136_{18}	0.007_1	0.594_7
	660	620_{56}	0.036_3	0.516_8
	680	876_{98}	0.06_1	0.49_1
	700	1150_{140}	0.084_9	0.47_1
	720	1209_{176}	0.13_2	0.42_2
OPLS-UA	550	9_2	0.0004_3	0.739_4
	600	22_3	0.0004_1	0.703_8
	660	95_{25}	0.004_1	0.660_7
	680	209_{40}	0.010_2	0.645_7
	700	245_{50}	0.011_2	0.625_9

Table C.9: Vapour-liquid coexistence simulation results for 1-hexadecanol.

Force Field	T [K]	P [kPa]	ρ_g [g/ml]	ρ_l [g/ml]
NERD	600	76_{12}	0.0050_8	0.61_1
	630	122_{19}	0.007_1	0.588_7
	660	223_{13}	0.014_1	0.557_8
	690	342_{40}	0.022_3	0.51_1
	710	409_{57}	0.025_3	0.487_8
	730	552_{87}	0.035_5	0.46_2
TraPPE-UA	600	66_{13}	0.0042_8	0.617_7
	630	180_{22}	0.012_2	0.588_5
	660	250_{18}	0.016_1	0.561_4
	690	450_{49}	0.032_6	0.52_1
	710	564_{41}	0.041_3	0.49_1
	730	758_{71}	0.069_6	0.470_9
OPLS-UA	600	9_2	0.0005_1	0.725_4
	630	13_4	0.0003_2	0.702_6
	660	19_3	0.0008_1	0.679_5
	690	47_8	0.0027_7	0.662_6
	710	57_{11}	0.002_1	0.649_4
	730	81_{19}	0.003_1	0.62_1

Table C.10: Vapour-liquid coexistence simulation results for 1-eicosanol.

C.4 Binary n-Alkane Mixtures

Only two binary n-alkane mixtures were briefly studied, the n-pentane + n-octane and n-hexane + n-hexadecane systems. Only the NERD force field was used for these simulations.

P [kPa]	x_1	y_1
4.03	0.000	0.000
9.32	0.042	0.501
17.87	0.161	0.776
38.00	0.368	0.938
58.94	0.551	0.964
79.55	0.768	0.984
101.54	1.000	1.000

Table C.11: P-x-y simulation results for n-pentane(1) + n-octane(2) at 308.7K.

P [kPa]	x_1	y_1
0.0003	0.000	0.000
0.52	0.041	0.804
2.96	0.144	0.922
7.42	0.354	1.000
11.53	0.563	1.000
16.12	0.775	1.000
18.28	0.871	1.000
22.55	1.000	1.000

Table C.12: P-x-y simulation results for n-hexane(1) + n-hexadecane(2) at 298.15K.

Appendix D

Literature Data

T [K]	P [kPa]	ρ_g [g/ml]	ρ_l [g/ml]
400	105.3	0.61004	0.003838
408	130.6	0.60242	0.004715
416	160.4	0.59461	0.005743
424	195.2	0.58661	0.006942
432	235.5	0.57841	0.008332
440	281.8	0.57001	0.009937
448	334.8	0.56133	0.011785
456	394.9	0.55234	0.013905
464	462.9	0.54305	0.016335
472	539.5	0.53334	0.019118
480	625.2	0.52317	0.022306
488	720.9	0.51245	0.025961
496	827.3	0.50110	0.030164
504	945.3	0.48898	0.035018
512	1076	0.47592	0.040666
520	1220	0.46170	0.047320
528	1379	0.44604	0.055237
536	1555	0.42855	0.064940
552	1961	0.38563	0.078888

Table D.1: Experimental vapour-liquid coexistence data for n-octane (Smith and Srivastava, 1986*a*).

T [K]	P [kPa]	ρ_l [g/ml]	T [K]	P [kPa]
443.15		0.6674	443.08	3.27
453.15		0.6599	453.08	4.81
463.15	6.83	0.6522	463.09	6.92
473.15	9.64	0.6447	473.13	9.77
483.15	13.37	0.637	483.11	13.53
493.15	18.32	0.629	493.13	18.41
503.15	24.38	0.621	503.12	24.69
513.15	32.21	0.612	513.13	32.60
523.15	41.94	0.604	523.13	42.46
533.15	53.93	0.596	533.13	54.64
543.15	68.68	0.587	543.13	69.44
553.15	86.14	0.578	553.13	87.29
			563.13	108.62
			573.13	133.94
			583.13	163.53

Table D.2: Experimental vapour-liquid coexistence data for n-hexadecane (Smith and Srivastava, 1986*a*; *DDB - Dortmund Data Bank*, 2004).

T [K]	P [kPa]
523.12	2.209
533.13	3.176
543.13	4.489
558.13	7.303
573.13	11.470
588.13	17.460

Table D.3: Experimental vapour-pressure data for n-tetracosane (*DDB - Dortmund Data Bank*, 2004).

T [K]	P [kPa]	V_l^{sat} [ml/mol]	V_g^{sat} [ml/mol]
340	16.41	166.34	169617
346	20.74	167.66	136203
352	26	169	110227
358	32.31	170.37	89874
364	39.84	171.77	73821
370	48.71	173.2	61090
376	59.07	174.66	50946
382	70.99	176.16	42834
388	84.53	177.7	36328
393	97.03	179.01	31895
405		182.27	
411		183.97	
417		185.72	
429		189.39	
435		191.32	
441		193.32	
453		197.55	
458		199.41	
464		201.74	
476		206.76	
482		209.48	
488		212.37	
500		218.75	
506		222.33	
512		226.23	
518		230.53	
523		234.49	

Table D.4: Experimental vapour-liquid coexistence data for 1-octene (Smith and Srivastava, 1986*a*).

1-decene		1-hexadecene	
T [K]	P [kPa]	T [K]	P [kPa]
359.92	6.40	461.30	6.94
364.46	7.69	470.41	9.54
368.28	9.00	480.13	13.07
371.75	10.33	492.96	19.44
374.99	11.72	498.55	22.89
379.37	13.85	506.35	28.55
384.36	16.65	521.84	43.18
389.43	19.95	530.59	53.78
394.14	23.48	539.74	67.06
400.42	28.99	556.55	98.14
413.21	43.36	557.92	101.09
420.42	53.71		
428.09	66.81	420.93	1.33
442.28	97.66	437.09	2.67
442.91	99.25	460.93	6.67
443.49	100.75	482.04	13.33
444.16	102.48	504.65	26.66
444.76	104.03	530.26	53.33
		558.15	101.32

Table D.5: Experimental vapour-pressure data for 1-decene (left-hand columns) and 1-hexadecene (right-hand columns) (*DDB - Dortmund Data Bank*, 2004).

T [K]	P [kPa]	V_l^{sat} [ml/mol]	V_g^{sat} [ml/mol]
397	8.27	175.17	397055
406	12.21	176.94	274481
414	16.91	178.57	201671
423	23.89	180.46	145553
431	31.89	182.19	110793
440	43.33	184.21	82953
448	56.04	186.07	65072
456	71.51	187.99	51689
465	92.68	190.24	40447
473	115.3	192.32	32894
482	145.5	194.76	26372
490	177.1	197.04	21875
499	218.5	199.73	17896
507	261.1	202.24	15089
515	309.5	204.89	12806
524	371.6	208.05	10724
532	434.1	211.05	9210
541	513.4	214.66	7804
549	592.6	218.12	6763
566		226.44	
583		236.56	
591		242.21	
600		249.48	
608		257	
617		267.06	
625		277.98	
633		291.55	

Table D.6: Experimental vapour-liquid coexistence data for 1-octanol (Smith and Srivastava, 1986*b*).

T [K]	P [kPa]	V_l^{sat} [ml/mol]	V_g^{sat} [ml/mol]
400	0.73	246.40	4572361
406	1.005	247.92	3358430
412	1.37	249.48	2501160
417	1.754	250.8	1976152
423	2.335	252.42	1506485
429	3.07	254.09	1161986
435	3.991	255.79	906241
440	4.926	257.24	742605
446	6.284	259.02	590109
452	7.939	260.85	473369
458	9.939	262.72	383123
463	11.91	264.33	323291
469	14.68	266.3	265642
475	17.96	268.33	219943
481	21.8	270.43	183426
486	25.49	272.22	158500
492	30.57		133817
498	36.42		113680
504	43.14		97143
509	49.45		85580
515	57.97		73867
521	67.6		64081
527	78.44		55859
532	88.47		49995
538	101.8		43939
544	116.6		38779
550	133.1		34360

Table D.7: Experimental vapour-liquid coexistence data for 1-dodecanol (Smith and Srivastava, 1986*b*).

T [K]	P [kPa]	V_l^{sat} [ml/mol]	V_g^{sat} [ml/mol]
499	7.06	352.18	587534
505	8.665	354.6	484569
511	10.56	357.06	402359
516	12.39	359.15	346309
522	14.92	361.71	290885
528	17.86	364.31	245759
534	21.27	366.96	208785
539	24.49	369.22	182995
545	28.87	371.97	156935
551	33.88	374.79	135231
556	38.56	377.18	119871
562	44.87	380.11	104136
573		385.64	

Table D.8: Experimental vapour-liquid coexistence data for 1-hexadecanol (Smith and Srivastava, 1986*b*).

T[K]	P[kPa]
600	52
630	104
660	189
690	318
710	434
730	577

Table D.9: Predicted vapour-pressure data for 1-eicosanol using the KDB Antoine Equation (*KDB - Korea Thermophysical Properties Data Bank*, 2003).

P [kPa]	x_1	y_1
3.39	0	0
5.7	0.022	0.3933
14	0.1503	0.8005
25.8	0.2528	0.9
29.8	0.2885	0.9248
39.5	0.3737	0.9455
47.2	0.4544	0.9597
58.3	0.5657	0.9721
67.7	0.6547	0.9855
77.2	0.7394	0.9927
86.35	0.8424	0.9974
95.2	0.9267	0.9993
100.6	1	1

Table D.10: Experimental P-x-y data for n-pentane(1) + n-octane(2) at 308.7K (*KDB - Korea Thermophysical Properties Data Bank*, 2003).

P [kPa]	x_1
1.972	0.108
3.858	0.2076
5.788	0.3061
7.574	0.3949
9.693	0.4984
9.707	0.4996
12.207	0.6196
14.025	0.7055
16.053	0.8013
17.989	0.894

Table D.11: Experimental P-x-y data for n-hexane(1) + n-hexadecane(2) at 298.15K (*KDB - Korea Thermophysical Properties Data Bank*, 2003).

Appendix E

Data Extraction BASH Script

This BASH script was originally created by T. McKnight for use on BIGMAC output files. It was modified to extract similar data from the TOWHEE output files. It has subsequently been added to the "Utils" section in the TOWHEE package. For more information on this script (and it's most recent form) see the TOWHEE internet site (Martin, 2004*a*).

```
#!/bin/bash
# script to process the output from a single component
# VLCC simulation run
# originally written 8-2004 by N. du Preez
# last modified 09-01-2004 by M.G. Martin
outputfile=$1
rm -rf $PWD/Plots
mkdir $PWD/Plots

# Get instantaneous energies, volumes, pressures, molecules:
#-----------------------------------------------------------
grep "B: 1" $outputfile | gawk '{ print $4 }' > Plots/energy_box1
grep "B: 2" $outputfile | gawk '{ print $3 }' > Plots/energy_box2
grep "B: 1" $outputfile | gawk '{ print $5 }' > Plots/volume_box1
grep "B: 2" $outputfile | gawk '{ print $4 }' > Plots/volume_box2
grep "B: 1" $outputfile | gawk '{ print $6 }' > Plots/pressure_box1
grep "B: 2" $outputfile | gawk '{ print $5 }' > Plots/pressure_box2
grep "B: 1" $outputfile | gawk '{ print $7 }' > Plots/molecules_box1
grep "B: 2" $outputfile | gawk '{ print $6 }' > Plots/molecules_box2
grep "1 Number Density" $outputfile | gawk '{ print $6 }' > Plots/density_box1
grep "2 Number Density" $outputfile | gawk '{ print $6 }' > Plots/density_box2
grep "DHvap" $outputfile | gawk '{ print $4 }' > Plots/DHvap
```

```
# Get block averaged values for energy, volume, pressure,
# mole fraction, specific density, number density, CB-mu:
#-------------------------------------------------------
grep "BA Box: 1 Specific density" $outputfile \
| gawk '{ print $7 }' > Plots/avgden_box1
grep "BA Box: 2 Specific density" $outputfile \
| gawk '{ print $7 }' > Plots/avgden_box2
grep "BA Box: 1 Pressure" $outputfile \
| gawk '{ print $6 }' > Plots/avgpres_box1
grep "BA Box: 2 Pressure" $outputfile \
| gawk '{ print $6 }' > Plots/avgpres_box2
grep "BA Box: 1 Total energy" $outputfile \
| gawk '{ print $6 }' > Plots/avgeng_box1
grep "BA Box: 2 Total energy" $outputfile \
| gawk '{ print $6 }' > Plots/avgeng_box2
grep "BA Box: 1 Chemical Potential" $outputfile \
| gawk '{ print $9 }' > Plots/avgchempot_box1
grep "BA Box: 2 Chemical Potential" $outputfile \
| gawk '{ print $9 }' > Plots/avgchempot_box2
grep "BA Box: 1 Number density" $outputfile \
| gawk '{ print $9 }' > Plots/avgnumden_box1
grep "BA Box: 2 Number density" $outputfile \
| gawk '{ print $9 }' > Plots/avgnumden_box2

# Check for any problems with the simulations
# (such as box vols too small...)
echo Number of volume moves failed: `grep "VOLN" $outputfile | wc -l`
# (such as problems in energy)
echo Number of problem warnings: `grep "problem" $outputfile | wc -l`
# (such as box empties of all molecules)
echo Number of times box 1 empties: `grep ^0 Plots/molecules_box1 | wc -l`
echo Number of times box 2 empties: `grep ^0 Plots/molecules_box2 | wc -l`
```

Appendix F

Abridged Output File

In this appendix the abridged output file from a standard TOWHEE simulation run has been shown. This file is generated by redirecting (`towhee > outfile`) the output data to a file of any user-specified name. The purpose of this appendix is to give an impression of the amount of data generated by a single simulation run. Also, many of the setting presented here are typical values for simulation settings. The output file shown here is from the production run of 1-eicosanol using the NERD force field at 660K. Typically BASH scripts (Appendix E) were then used to extract relevant information such as change in pressure, densities, volumes and block averages during the simulation.

F.1 Simulation Settings Printouts

This section is the first part of the output file, it lists all the relevant simulation details. After these settings are printed out an initial system energy balance is performed for simulation consistency checks that are performed during the simulation.

```
 Reading from towhee_input
inputformat:Towhee
 RANLUX LUXURY LEVEL SET BY RLUXGO : 3     P= 223
 RANLUX INITIALIZED BY RLUXGO FROM SEEDS     1302002           0           0
 Testing random number generator
    0.354084  0.719681  0.403250  0.706761  0.174308
ensemble: nvt
temperature:            660.00000
nmolty:                   1
nmolectyp:       100
numboxes         2
```

```
stepstyle:cycles
nstep:                50000
printfreq:                50
blocksize:              5000
moviefreq:             10000
backupfreq:             10000
loutpdb: F
loutdft: F
loutlammps: F
pressurefreq:                25
trmaxdispfreq:             5000
volmaxdispfreq:             5000
ffnumber:                  1
ff_filename:
/home/nicholas/towhee-3.17.5/ForceFields/towhee_ff_NERDv3
potentype: 0
mixrule: 0
lshift: F
ltailc: T
rmin:   0.00000
rcut:  14.00000
rcutin:   5.00000
coulombstyle: ewald_fixed_kmax
kalp:   5.00000
kmax:    5
dielect:    1.00000
 Setting up force field parameters from files
 opening forcefield file:    1
 Lorentz-Berthelot Mixing rules
 Arithmetic mean of sigma
 Geometric mean of epsilon
nhrdfld:    0
nljfld:    0
numbfld:   0
 No solvation model used
linit: F
initstyle Box:    1
    0
initstyle Box:    2
    0
Box idim hmatrix:    1 1   40.00000    0.00000    0.00000
```

```
Box idim hmatrix:    1 2    0.00000   40.00000    0.00000
Box idim hmatrix:    1 3    0.00000    0.00000   40.00000
Box idim hinverse:    1 1    0.02500    0.00000    0.00000
Box idim hinverse:    1 2    0.00000    0.02500    0.00000
Box idim hinverse:    1 3    0.00000    0.00000    0.02500
Box idim hmatrix:    2 1  100.00000    0.00000    0.00000
Box idim hmatrix:    2 2    0.00000  100.00000    0.00000
Box idim hmatrix:    2 3    0.00000    0.00000  100.00000
Box idim hinverse:    2 1    0.01000    0.00000    0.00000
Box idim hinverse:    2 2    0.00000    0.01000    0.00000
Box idim hinverse:    2 3    0.00000    0.00000    0.01000
Box:  1 initmol:        60
Box:  2 initmol:        40
Box:  1 inix,iniy,iniz:         6          6          6
Box:  2 inix,iniy,iniz:         4          4          4
Box:  1 inimix:         1
Box:  2 inimix:         1
itest:  1 pairbox:  1   2
pmvol: 0.006000
pmvlpr: 1.000000
rmvol:   0.10000
tavol:   0.50000
pmcell: 0.000000
pmcellpr: 1.000000
pairbox: 1  2 pmcellpt: 0.500000
rmcell:   1.00000
tacell:   0.50000
pm2boxrbswap: 0.000000
pm2rbswmt: 1.000000
pm2rbswpr: 1.000000
pm2boxcbswap: 0.334000
pm2cbswmt: 1.000000
pm2cbswpr: 1.000000
pm1boxcbswap: 0.000000
pm1cbswmt: 1.000000
pmavb1: 0.000000
pmavb1in: 0.500000
pmavb1mt: 1.000000
moltyp:  1 pmavb1ct: 1.000000
avb1rad:   8.00000
pmavb2: 0.000000
```

```
pmavb2in: 0.500000
pmavb2mt: 1.000000
moltyp:  1 pmavb2ct: 1.000000
avb2rad:   8.00000
pmavb3: 0.000000
pmavb3mt: 1.000000
moltyp:  1 pmavb3ct: 1.000000
avb3rad:   8.00000
pmcb: 0.556000
pmcbmt: 1.000000
pmall: 0.000000
pmback: 0.000000
pmbkmt: 1.000000
pmpivot 0.000000
pmpivmt: 1.000000
pmconro 0.000000
pmcrmt: 1.000000
pmcrbac 0.000000
pmcrbmt: 1.000000
pmplane: 0.000000
pmplanebox: 0.500000  1.000000
planewidth:   3.00000
pmrow: 0.000000
pmrowbox: 0.500000  1.000000
rowwidth:   3.00000
pmtraat: 0.000000
pmtamt: 1.000000
rmtraa:   0.50000
tatraa:   0.50000
pmtracm:   0.77800
pmtcmt: 1.000000
rmtrac:   0.80000
tatrac: 0.500000
pmrotate: 1.000000
pmromt: 1.000000
rmrot:   0.05000
tarot:   0.50000
tor_cbstyle:  1
 Use gaussian distribution to generate torsions in config-bias
    and correct this in the rosenbluth weights
bend_cbstyle:   1
```

```
 Use gaussian distribution to generate angle in config-bias
    and correct this in the rosenbluth weights
vib_cbstyle:   1
 Use gaussian distribution to generate vibrations in config-bias
    and correct this in the rosenbluth weights
sdevtor:  10.00000
sdevbena:   5.00000
sdevbenb:  10.00000
sdevvib:   0.10000
vibrang:   0.85000    1.15000
cdform: 0
 Coupled-decoupled form from M.G. Martin;
 J.I. Siepmann; J. Phys. Chem. B 103 2977-2980 (1999)
nch_nb_one:        20
nch_nb:        20
nch_tor_out:         1
nch_tor_in:        20
nch_tor_in_con:        10
nch_bend_a:        20
nch_bend_b:        20
nch_vib:        50
inpstyle:   2
nunit:         22
nmaxcbmc:         11
   Building the input file for molecule type:     1
   using the NERDv3     force field
unit:     1 name:CH3sp3gen  charge:   0.00000
unit:     2 name:CH2sp3     charge:   0.00000
unit:     3 name:CH2sp3     charge:   0.00000
unit:     4 name:CH2sp3     charge:   0.00000
unit:     5 name:CH2sp3     charge:   0.00000
unit:     6 name:CH2sp3     charge:   0.00000
unit:     7 name:CH2sp3     charge:   0.00000
unit:     8 name:CH2sp3     charge:   0.00000
unit:     9 name:CH2sp3     charge:   0.00000
unit:    10 name:CH2sp3     charge:   0.00000
unit:    11 name:CH2sp3     charge:   0.00000
unit:    12 name:CH2sp3     charge:   0.00000
unit:    13 name:CH2sp3     charge:   0.00000
unit:    14 name:CH2sp3     charge:   0.00000
unit:    15 name:CH2sp3     charge:   0.00000
```

```
unit:   16 name:CH2sp3     charge:   0.00000
unit:   17 name:CH2sp3     charge:   0.00000
unit:   18 name:CH2sp3     charge:   0.00000
unit:   19 name:CH2sp3     charge:   0.00000
unit:   20 name:CH2sp3     charge:   0.29000
unit:   21 name:Osp3       charge:  -0.71000
unit:   22 name:H_o        charge:   0.42000
Default total charge on molecule   1 is   0.00000
Total charge in the simulation system:   0.00000
Bond Types
Type:   1 Style: Standard Harmonic Length: 1.5400 Constant:   168380.4
Type:   4 Style: Standard Harmonic Length: 1.4280 Constant:   198448.3
Type:   5 Style: Standard Harmonic Length: 0.9610 Constant:   312706.4
Angle Types
Type:   1 Style: Standard Harmonic  Angle:    114.000 Constant:    45703.2
Type:   5 Style: Standard Harmonic  Angle:    108.000 Constant:    60135.8
Type:   6 Style: Standard Harmonic  Angle:    107.500 Constant:    27662.5
Torsion Types
Type:   1 Style: Old UA OPLS Cosine Series
          k0:       0.0 k1:     355.0 k2:     -68.2 k3:     791.3
Type:   2 Style: Old UA OPLS Cosine Series
          k0:       0.0 k1:     355.0 k2:     -68.2 k3:     791.3
Type:   6 Style: Old UA OPLS Cosine Series
          k0:     232.0 k1:     356.4 k2:    -472.2 k3:    1119.8
Type:   7 Style: Old UA OPLS Cosine Series
          k0:       0.0 k1:     359.3 k2:      59.1 k3:     220.8
Improper Torsion Types
   No Improper Types
 Verifying input structures are consistent
 Canonical Gibbs ensemble
 3-dimensional periodic box
 Additional Center-of-Mass cutoff
 Dual Cutoff Configurational-bias Monte Carlo
 Coupled-decoupled Configurational-bias MC
 Coulombic inter- and intra-molecular interactions
     with an Ewald sum
     including the real-space terms up to half the shortest box length
Molecular mass for molecule type     1 is   298.5508 g/mol
 Reading in initial conformation from towhee_inital
Initial version:    3
 new maximum displacements read from towhee_initial
```

```
box:     1
molecule type:     1
 Max disp. for Atom translate:      0.500000  0.500000  0.500000
 Max disp. for COM translate:       0.725399  0.715976  0.708525
 Max disp. for Rotate:              0.150554  0.148852  0.148432
box:     2
molecule type:     1
 Max disp. for Atom translate:      0.500000  0.500000  0.500000
 Max disp. for COM translate:      28.000000 28.000000 28.000000
 Max disp. for Rotate:              3.141500  3.141500  3.141500
 Max disp. for 3D Volume:              0.2438E-01  0.1000E+00
 Max disp. for unit cell perturbation
    Boxes   1 and   2 idim 1 rmcell:   0.1000E+01  0.1000E+01  0.1000E+01
    Boxes   1 and   2 idim 2 rmcell:   0.1000E+01  0.1000E+01  0.1000E+01
    Boxes   1 and   2 idim 3 rmcell:   0.1000E+01  0.1000E+01  0.1000E+01
 new box dimensions read from towhee_initial
 Box  1 hmatrix(1,x):        40.82612        0.00000        0.00000
 Box  1 hmatrix(2,x):         0.00000       40.82612        0.00000
 Box  1 hmatrix(3,x):         0.00000        0.00000       40.82612
 Box  2 hmatrix(1,x):        99.86489        0.00000        0.00000
 Box  2 hmatrix(2,x):         0.00000       99.86489        0.00000
 Box  2 hmatrix(3,x):         0.00000        0.00000       99.86489
Box:     1 Initial calp:    0.12247 Box:     1 Initial kmax:     5
Box:     2 Initial calp:    0.05007 Box:     2 Initial kmax:     5
 Nonbonded Force Field
 Lennard-Jones 12-6 potential
     with tail corrections
 u(r) = 4*epsilon[(sigma/r)^12 - (sigma/r)^6] - S
Num. Atom(i)     Num. Atom(j)      sigma epsilon   shift  1-4sig  1-4eps
   4 CH3sp3gen      4 CH3sp3gen    3.910 103.999   0.000   0.000   0.000
   4 CH3sp3gen     12 CH2sp3       3.920  69.015   0.000   0.000   0.000
   4 CH3sp3gen     19 H_o          2.445  20.131   0.000   0.000   0.000
   4 CH3sp3gen     20 Osp3         3.445 105.977   0.000   0.000   0.000
  12 CH2sp3        12 CH2sp3       3.930  45.799   0.000   0.000   0.000
  12 CH2sp3        19 H_o          2.455  13.359   0.000   0.000   0.000
  12 CH2sp3        20 Osp3         3.455  70.328   0.000   0.000   0.000
  19 H_o           19 H_o          0.980   3.897   0.000   0.000   0.000
  19 H_o           20 Osp3         1.980  20.514   0.000   0.000   0.000
  20 Osp3          20 Osp3         2.980 107.992   0.000   0.000   0.000
Number of MC cycles:                50000
Number of molecules:                  100
```

```
Temperature [K]:                660.00000

Initial Energies from engtotal for Box     1
Total molecules in this box          68
Molecules of type   1 :          68
 total vibration       485567.779 [K]       964.92796 [kcal/mol]
    regular            485567.779 [K]       964.92796 [kcal/mol]
    bond-bond(1-2)          0.000 [K]         0.00000 [kcal/mol]
 total angle           420273.493 [K]       835.17412 [kcal/mol]
    regular            420273.493 [K]       835.17412 [kcal/mol]
    angle-angle             0.000 [K]         0.00000 [kcal/mol]
 total torsion         668791.784 [K]      1329.03359 [kcal/mol]
    regular            668791.784 [K]      1329.03359 [kcal/mol]
    improper                0.000 [K]         0.00000 [kcal/mol]
 total nonbond        -434338.889 [K]      -863.12510 [kcal/mol]
    intramolecular     -71455.960 [K]      -141.99841 [kcal/mol]
    2-body nonbond    -347139.047 [K]      -689.84020 [kcal/mol]
    3-body nonbond          0.000 [K]         0.00000 [kcal/mol]
    tail correct.      -15743.882 [K]       -31.28649 [kcal/mol]
 total coulombic       -19766.512 [K]       -39.28032 [kcal/mol]
    real space         -19533.705 [K]       -38.81769 [kcal/mol]
    self              -600311.130 [K]     -1192.94776 [kcal/mol]
    correction         598461.583 [K]      1189.27231 [kcal/mol]
    recip sum            1616.740 [K]         3.21281 [kcal/mol]
 external field          0.000 [K]         0.00000 [kcal/mol]
 solvation               0.000 [K]         0.00000 [kcal/mol]
 total energy         1120527.654 [K]      2226.73025 [kcal/mol]

Initial Energies from engtotal for Box     2
Total molecules in this box          32
Molecules of type   1 :          32
 total vibration       228493.595 [K]       454.06608 [kcal/mol]
    regular            228493.595 [K]       454.06608 [kcal/mol]
    bond-bond(1-2)          0.000 [K]         0.00000 [kcal/mol]
 total angle           212193.223 [K]       421.67372 [kcal/mol]
    regular            212193.223 [K]       421.67372 [kcal/mol]
    angle-angle             0.000 [K]         0.00000 [kcal/mol]
 total torsion         341793.267 [K]       679.21698 [kcal/mol]
    regular            341793.267 [K]       679.21698 [kcal/mol]
    improper                0.000 [K]         0.00000 [kcal/mol]
 total nonbond         -50390.218 [K]      -100.13624 [kcal/mol]
```

```
   intramolecular      -37925.567 [K]        -75.36629 [kcal/mol]
   2-body nonbond      -12226.436 [K]        -24.29657 [kcal/mol]
   3-body nonbond           0.000 [K]          0.00000 [kcal/mol]
   tail correct.         -238.215 [K]         -0.47339 [kcal/mol]
total coulombic         -1368.835 [K]         -2.72017 [kcal/mol]
   real space           -1358.444 [K]         -2.69952 [kcal/mol]
   self               -115489.568 [K]       -229.50269 [kcal/mol]
   correction          115432.775 [K]        229.38983 [kcal/mol]
   recip sum               46.403 [K]          0.09221 [kcal/mol]
external field              0.000 [K]          0.00000 [kcal/mol]
solvation                   0.000 [K]          0.00000 [kcal/mol]
total energy           730721.031 [K]       1452.10037 [kcal/mol]
```

F.2 Runtime Simulation Printouts

Here the runtime printouts have been shown for the first 500 cycles of the simulation. These printouts actually continue until the end of the simulation which is after 50000 cycles.

```
 initial pressure in box  1 =      -19668.39
 initial pressure in box  2 =         204.43
 +++++ start of markov chain +++++
Cycle  Box   Energy [K]  Volume [A^3] Press. [kPa] Molecules
    50 B: 1  0.1141E+07  0.6766E+05      -7150.3    68
 1 Number Density [/nm3] =   0.0010049874
       B: 2  0.7226E+06  0.9963E+06        152.0    32
 2 Number Density [/nm3] =   3.21176322E-05
 DHvap [kJ/mol] =  -51.0129565
   100 B: 1  0.1206E+07  0.6818E+05     -21793.0    71
 1 Number Density [/nm3] =   0.00104133445
       B: 2  0.6622E+06  0.9958E+06        262.1    29
 2 Number Density [/nm3] =   2.912178E-05
 DHvap [kJ/mol] =  -53.8441098
   150 B: 1  0.1173E+07  0.6803E+05     -41005.1    71
 1 Number Density [/nm3] =   0.00104362894
       B: 2  0.6634E+06  0.9960E+06        295.5    29
 2 Number Density [/nm3] =   2.91173969E-05
 DHvap [kJ/mol] =  -58.7642555
   200 B: 1  0.1144E+07  0.6753E+05       -320.6    69
 1 Number Density [/nm3] =   0.00102184136
       B: 2  0.7099E+06  0.9965E+06        195.0    31
```

```
 2 Number Density [/nm3] =   3.11096666E-05
 DHvap [kJ/mol] =  -56.1683812
   250 B: 1  0.1132E+07  0.6723E+05      19631.1    70
 1 Number Density [/nm3] =   0.0010411958
       B: 2  0.6909E+06  0.9968E+06        411.0    30
 2 Number Density [/nm3] =   3.00972259E-05
 DHvap [kJ/mol] =  -65.0253957
   300 B: 1  0.1167E+07  0.6839E+05     -13323.4    70
 1 Number Density [/nm3] =   0.00102359204
       B: 2  0.6750E+06  0.9956E+06        454.9    30
 2 Number Density [/nm3] =   3.01321785E-05
 DHvap [kJ/mol] =  -57.2373231
   350 B: 1  0.1186E+07  0.6824E+05      -4694.5    71
 1 Number Density [/nm3] =   0.00104048787
       B: 2  0.6872E+06  0.9958E+06        151.2    29
 2 Number Density [/nm3] =   2.91234024E-05
 DHvap [kJ/mol] =  -61.1835719
   400 B: 1  0.1109E+07  0.6744E+05     -17072.7    67
 1 Number Density [/nm3] =   0.000993422041
       B: 2  0.7754E+06  0.9966E+06        125.2    33
 2 Number Density [/nm3] =   3.3114033E-05
 DHvap [kJ/mol] =  -60.0177138
   450 B: 1  0.1165E+07  0.6585E+05      16520.8    71
 1 Number Density [/nm3] =   0.00107823422
       B: 2  0.6782E+06  0.9982E+06        127.2    29
 2 Number Density [/nm3] =   2.9053703E-05
 DHvap [kJ/mol] =  -60.594038
   500 B: 1  0.1137E+07  0.6586E+05     -13245.4    72
 1 Number Density [/nm3] =   0.00109328552
       B: 2  0.6446E+06  0.9981E+06        160.4    28
 2 Number Density [/nm3] =   2.80520799E-05
 DHvap [kJ/mol] =  -63.4799132
```

As per the simulation settings, every 5000 cycles the maximum translations, rotations and volume moves were updated. The printout showing this update is as follows:

```
            Updating maximum translational/rotational displacements
 Box:  1          X Atmpt Y Atmpt Z Atmpt  X Disp. Y Disp. Z Disp.
 Molecule Type:    1
 Translate COM      27200   27143   27273  0.7353  0.7230  0.7373
 Translate Atom         0       0       0  0.5000  0.5000  0.5000
 Rotate             26997   27126   27059  0.1565  0.1532  0.1531
```

```
Box:  2          X Atmpt Y Atmpt Z Atmpt  X Disp. Y Disp. Z Disp.
Molecule Type:    1
Translate COM       9804    9886    9902 28.0000 28.0000 28.0000
Translate Atom         0       0       0  0.5000  0.5000  0.5000
Rotate              9952    9775    9844  3.1415  3.1415  3.1415
Updating 3D volume maximum displacements
Boxes  1 and  2 Tries:     2952 Accepted:     1455 Max Disp.: 0.240E-01
```

In this particular simulation the block averages were also set to be determined after 5000 cycles. This printout shows the results of the first block average determined.

```
Block Averages (BA) for block     1
BA Box: 1 Specific density [g/ml] 0.53684818E+00
BA Box: 1 Pressure [kPa] 0.22849533E+03
BA Box: 1 Total energy    0.11942287E+07
BA Box: 1 Inter vdw       -.42322221E+06
BA Box: 1 Bond bending    0.46743521E+06
BA Box: 1 Torsion         0.73609449E+06
BA Box: 1 Intra vdw       -.79190638E+05
BA Box: 1 External pot    0.00000000E+00
BA Box: 1 Vibration       0.51242934E+06
BA Box: 1 Coulomb         -.19317494E+05
BA Box: 1 Tail vdw        -.18436106E+05
BA Box: 1 Solvation       0.00000000E+00
BA Box: 1 Chemical Potential [K] Type   1 0.32361238E+11
BA Box: 1 Number density [nm-3] Type   1 0.10828723E-02
BA Box: 1 Mol Fraction Type   1 0.10000000E+01
BA Box: 1 Stress Tensor S_xx     [kPa] 0.91265412E+04
BA Box: 1 Stress Tensor S_yy     [kPa] 0.10210100E+05
BA Box: 1 Stress Tensor S_zz     [kPa] 0.38660998E+04
BA Box: 1 Stress Tensor S_xy     [kPa] 0.93848939E+04
BA Box: 1 Stress Tensor S_xz     [kPa] 0.10961248E+05
BA Box: 1 Stress Tensor S_yz     [kPa] 0.11934392E+05
BA Box: 1 Stress Tensor P_tail   [kPa] -.75057516E+04
BA Box: 2 Specific density [g/ml] 0.13221083E-01
BA Box: 2 Pressure [kPa] 0.20177861E+03
BA Box: 2 Total energy    0.60995497E+06
BA Box: 2 Inter vdw       -.75172279E+04
BA Box: 2 Bond bending    0.17526781E+06
BA Box: 2 Torsion         0.28615372E+06
BA Box: 2 Intra vdw       -.30203662E+05
BA Box: 2 External pot    0.00000000E+00
```

```
BA Box: 2 Vibration        0.18673217E+06
BA Box: 2 Coulomb          -.47784472E+03
BA Box: 2 Tail vdw         -.16571449E+03
BA Box: 2 Solvation        0.00000000E+00
BA Box: 2 Chemical Potential [K] Type    1 0.20610748E+11
BA Box: 2 Number density [nm-3] Type    1 0.26668143E-04
BA Box: 2 Mol Fraction Type    1 0.10000000E+01
BA Box: 2 Stress Tensor S_xx      [kPa] 0.19595729E+03
BA Box: 2 Stress Tensor S_yy      [kPa] 0.21014361E+03
BA Box: 2 Stress Tensor S_zz      [kPa] 0.21298914E+03
BA Box: 2 Stress Tensor S_xy      [kPa] 0.24341171E+03
BA Box: 2 Stress Tensor S_xz      [kPa] 0.23250474E+03
BA Box: 2 Stress Tensor S_yz      [kPa] 0.24990605E+03
BA Box: 2 Stress Tensor P_tail    [kPa] -.45847355E+01
```

F.3 Final Results Printout

Finally, at the end of the simulation the final results and the block averages are computed and printed to screen. Also in these printouts the effectiveness of the CBMC technique is shown, and the final system energies are recomputed to check whether the simulation was consistent with regard to internal energy.

```
 +++++ end of markov chain +++++
 Final hmatrix (general box dimensions)
Box:      1
  hmatrix(1,x)       39.29264        0.00000        0.00000
  hmatrix(2,x)        0.00000       39.29264        0.00000
  hmatrix(3,x)        0.00000        0.00000       39.29264
Box:      2
  hmatrix(1,x)      100.11106        0.00000        0.00000
  hmatrix(2,x)        0.00000      100.11106        0.00000
  hmatrix(3,x)        0.00000        0.00000      100.11106


 * 3D Volume Change Moves *
 Box  1 and  2 Tries:     29638 Accepted:     14814
              Acp. Ratio:  0.500 Max Disp.: 0.242E-01
 * Configurational-Bias SWAP Moves *
Molecule type:      1
  From box  2 to box  1 Attempted:  820880 Grown:  820880 Accepted:     3322
  From box  1 to box  2 Attempted:  819604 Grown:  819604 Accepted:     3322
 * Configurational-Bias REGROWTH Moves *
```

```
 Molecule type:    1 Box:    1
     Length  Attempts  Regrown  Accepted  %Regrown   %Accep.
        1     78705     78705     45978    100.00     58.42
        2     78360     78360     30130    100.00     38.45
        3     79179     79179     21406    100.00     27.03
        4     78803     78803     15404    100.00     19.55
        5     79101     79101     11478    100.00     14.51
        6     79232     79232      8465    100.00     10.68
        7     78987     78987      6699    100.00      8.48
        8     79027     79027      4965    100.00      6.28
        9     79081     79081      3797    100.00      4.80
       10     39502     39502      1476    100.00      3.74
       11     39715     39715      1106    100.00      2.78

 Molecule type:    1 Box:    2
     Length  Attempts  Regrown  Accepted  %Regrown   %Accep.
        1     31944     31944     22555    100.00     70.61
        2     32000     32000     19728    100.00     61.65
        3     32351     32351     16970    100.00     52.46
        4     32025     32025     14719    100.00     45.96
        5     32116     32116     13091    100.00     40.76
        6     32155     32155     11350    100.00     35.30
        7     31811     31811     10113    100.00     31.79
        8     31969     31969      8926    100.00     27.92
        9     32043     32043      8055    100.00     25.14
       10     15932     15932      3661    100.00     22.98
       11     15936     15936      3302    100.00     20.72

Final Energies from engtotal for Box     1
Total molecules in this box         68
Molecules of type   1 :         68
 total vibration       475020.469 [K]       943.96818 [kcal/mol]
    regular            475020.469 [K]       943.96818 [kcal/mol]
    bond-bond(1-2)          0.000 [K]         0.00000 [kcal/mol]
 total angle           435554.746 [K]       865.54127 [kcal/mol]
    regular            435554.746 [K]       865.54127 [kcal/mol]
    angle-angle             0.000 [K]         0.00000 [kcal/mol]
 total torsion         655989.476 [K]      1303.59264 [kcal/mol]
    regular            655989.476 [K]      1303.59264 [kcal/mol]
    improper                0.000 [K]         0.00000 [kcal/mol]
 total nonbond        -480441.952 [K]      -954.74183 [kcal/mol]
```

```
    intramolecular     -72471.255 [K]      -144.01602 [kcal/mol]
    2-body nonbond    -390310.623 [K]      -775.63143 [kcal/mol]
    3-body nonbond          0.000 [K]         0.00000 [kcal/mol]
    tail correct.      -17660.074 [K]       -35.09438 [kcal/mol]
 total coulombic       -18646.812 [K]       -37.05524 [kcal/mol]
    real space         -18351.678 [K]       -36.46874 [kcal/mol]
    self              -623739.568 [K]     -1239.50512 [kcal/mol]
    correction         621585.201 [K]      1235.22393 [kcal/mol]
    recip sum            1859.233 [K]         3.69470 [kcal/mol]
 external field             0.000 [K]         0.00000 [kcal/mol]
 solvation                  0.000 [K]         0.00000 [kcal/mol]
 total energy         1067475.927 [K]      2121.30502 [kcal/mol]

Final Energies from engtotal for Box     2
Total molecules in this box         32
Molecules of type   1 :          32
 total vibration       226427.674 [K]       449.96065 [kcal/mol]
    regular            226427.674 [K]       449.96065 [kcal/mol]
    bond-bond(1-2)          0.000 [K]         0.00000 [kcal/mol]
 total angle           208737.856 [K]       414.80716 [kcal/mol]
    regular            208737.856 [K]       414.80716 [kcal/mol]
    angle-angle             0.000 [K]         0.00000 [kcal/mol]
 total torsion         308835.169 [K]       613.72212 [kcal/mol]
    regular            308835.169 [K]       613.72212 [kcal/mol]
    improper                0.000 [K]         0.00000 [kcal/mol]
 total nonbond         -46862.872 [K]       -93.12664 [kcal/mol]
    intramolecular     -35376.151 [K]       -70.30004 [kcal/mol]
    2-body nonbond     -11250.259 [K]       -22.35669 [kcal/mol]
    3-body nonbond          0.000 [K]         0.00000 [kcal/mol]
    tail correct.        -236.462 [K]        -0.46990 [kcal/mol]
 total coulombic         -187.580 [K]        -0.37276 [kcal/mol]
    real space           -193.958 [K]        -0.38544 [kcal/mol]
    self              -115205.577 [K]      -228.93834 [kcal/mol]
    correction         115145.794 [K]       228.81954 [kcal/mol]
    recip sum              66.161 [K]         0.13148 [kcal/mol]
 external field             0.000 [K]         0.00000 [kcal/mol]
 solvation                  0.000 [K]         0.00000 [kcal/mol]
 total energy          696950.247 [K]      1384.99053 [kcal/mol]

 Averages           Type    Units       Box  1       Box  2
 Pressure                     kPa      1298.12       222.70
```

S_xx		kPa	10187.64	222.43
S_yy		kPa	10338.74	229.73
S_zz		kPa	7642.11	232.06
S_xy		kPa	10873.40	265.75
S_xz		kPa	10200.54	264.63
S_yz		kPa	11893.02	264.45
P_tail		kPa	-8091.38	-5.38
CB Chem. Potential	1	K	-15372.602	-15848.566
Volume		nm^3	63.365	1000.635
Molecule Number	1		71.154	28.846
Molar Volume		ml/mol	535.869	20894.435
Specific Density		g/ml	0.557134	0.014289
Number Density	1	/nm^3	1.12379	0.02882
Mole Fraction	1		1.0000000	1.0000000
Radius of Gyration	1	A	5.7177740	5.6224494

Results for 10 blocks:			Average	Std Dev	Std Error
Specific Density	box	1 =	0.55713E+00	0.88196E-02	0.29399E-02
Specific Density	box	2 =	0.14289E-01	0.10294E-02	0.34312E-03
Pressure	box	1 =	1298.12	790.82	263.61
S_xx	box	1 =	10187.64	1709.16	569.72
S_yy	box	1 =	10338.74	2264.60	754.87
S_zz	box	1 =	7642.11	2824.90	941.63
S_xy	box	1 =	10873.40	2104.43	701.48
S_xz	box	1 =	10200.54	2218.56	739.52
S_yz	box	1 =	11893.02	2118.78	706.26
P_tail	box	1 =	-8091.38	255.51	85.17
Pressure	box	2 =	222.70	12.63	4.21
S_xx	box	2 =	222.43	18.72	6.24
S_yy	box	2 =	229.73	20.44	6.81
S_zz	box	2 =	232.06	16.25	5.42
S_xy	box	2 =	265.75	20.21	6.74
S_xz	box	2 =	264.63	25.73	8.58
S_yz	box	2 =	264.45	17.98	5.99
P_tail	box	2 =	-5.38	0.76	0.25
Total energy	box	1 =	0.11440E+07	0.45729E+05	0.15243E+05
Inter vdw	box	1 =	-0.42839E+06	0.15028E+05	0.50093E+04
Bond bending	box	1 =	0.45689E+06	0.18521E+05	0.61735E+04
Torsion	box	1 =	0.70837E+06	0.30972E+05	0.10324E+05
Intra vdw	box	1 =	-0.76465E+05	0.32769E+04	0.10923E+04
External pot	box	1 =	0.00000E+00	0.00000E+00	0.00000E+00

```
 Vibration            box  1 =  0.50507E+06 0.15579E+05 0.51930E+04
 Coulomb              box  1 = -0.21477E+05 0.15969E+04 0.53229E+03
 Tail vdw             box  1 = -0.18539E+05 0.66378E+03 0.22126E+03
 Solvation            box  1 =  0.00000E+00 0.00000E+00 0.00000E+00
 Total energy         box  2 =  0.64561E+06 0.45304E+05 0.15101E+05
 Inter vdw            box  2 = -0.82943E+04 0.12342E+04 0.41141E+03
 Bond bending         box  2 =  0.18658E+06 0.13051E+05 0.43502E+04
 Torsion              box  2 =  0.29764E+06 0.20794E+05 0.69315E+04
 Intra vdw            box  2 = -0.31859E+05 0.21194E+04 0.70647E+03
 External pot         box  2 =  0.00000E+00 0.00000E+00 0.00000E+00
 Vibration            box  2 =  0.20191E+06 0.15185E+05 0.50617E+04
 Coulomb              box  2 = -0.37090E+03 0.16189E+03 0.53963E+02
 Tail vdw             box  2 = -0.19517E+03 0.27888E+02 0.92961E+01
 Solvation            box  2 =  0.00000E+00 0.00000E+00 0.00000E+00
 CB Chem. Potential  Type  1 box  1 =   -15372.602     427.873     142.624
 CB Chem. Potential  Type  1 box  2 =   -15848.566     364.656     121.552
 Number Density nm-3 Type  1 box  1 =  0.11238E+01 0.17790E-01 0.59300E-02
 Number Density nm-3 Type  1 box  2 =  0.28821E-01 0.20763E-02 0.69210E-03
 Mole Fraction       Type  1 box  1 =    1.0000000   0.0000000   0.0000000
 Mole Fraction       Type  1 box  2 =    1.0000000   0.0000000   0.0000000
 Molarity            Type  1 box  1 =  0.18668E+01
 Molarity            Type  1 box  2 =  0.47876E-01

 -----block averages ------
Box:    1
Block Energy         Density        Pressure       Mol fracs
    1 0.11942287E+07 0.53684818E+00 0.22849533E+03 1.00000000
    2 0.10770700E+07 0.55102112E+00 0.18703109E+04 1.00000000
    3 0.11185371E+07 0.56878426E+00 -.21895026E+03 1.00000000
    4 0.12229188E+07 0.56982293E+00 0.14492152E+04 1.00000000
    5 0.11712654E+07 0.56144781E+00 0.70570258E+03 1.00000000
    6 0.11752026E+07 0.55782415E+00 0.14573140E+04 1.00000000
    7 0.11225184E+07 0.55557489E+00 0.12514197E+04 1.00000000
    8 0.10753643E+07 0.55771037E+00 0.18253595E+04 1.00000000
    9 0.11334124E+07 0.55820769E+00 0.24823541E+04 1.00000000
   10 0.11495708E+07 0.55409409E+00 0.19299814E+04 1.00000000
Box:    2
Block Energy         Density        Pressure       Mol fracs
    1 0.60995497E+06 0.13221083E-01 0.20177861E+03 1.00000000
    2 0.72074630E+06 0.15820076E-01 0.23761111E+03 1.00000000
    3 0.66470394E+06 0.14593902E-01 0.23522420E+03 1.00000000
```

```
    4 0.55368673E+06 0.12235285E-01 0.20669758E+03 1.00000000
    5 0.61174046E+06 0.13520747E-01 0.21572559E+03 1.00000000
    6 0.64141514E+06 0.14123135E-01 0.23045731E+03 1.00000000
    7 0.65588074E+06 0.14676803E-01 0.21939458E+03 1.00000000
    8 0.70355541E+06 0.15617095E-01 0.23933353E+03 1.00000000
    9 0.65615683E+06 0.14806827E-01 0.22791652E+03 1.00000000
   10 0.63823708E+06 0.14270367E-01 0.21281996E+03 1.00000000

real    5911m1.097s
user    5905m31.820s
sys 0m0.250s
  Source code :  /home/nicholas/towhee-3.17.5/Source
  Sun Nov 14 11:53:57 GMT 2004
```

The last five lines of this printout are part of the grind-engine printouts which gets tagged onto the end of this output file. It shows how much time the simulation took, the version of TOWHEE used, and the date of the simulation.

References

Alder, B. J. and Wainwright, T. E. (1957), 'Phase transition for a hard sphere system', *Journal of Chemical Physics* **27**, 1208–1209.

Allen, M. P. and Tildesley, D. J. (1987), *Computer Simulation of Liquids*, Clarendon Press, Oxford.

Bernstein, H. J. (2001), RasMol V2.7.2.1 – Molecular Graphics Visualisation Tool.

Chase, J. D. (1987), 'Two Models of Vapor Pressure along the Saturation Curve', *Industrial and Engineering Chemistry Research* **26**(1), 107–112.

Chen, B., Martin, M. G. and Siepmann, J. I. (1998), 'Thermodynamic Properties of the Williams, OPLS-AA, and MMFF94 All-Atom Force Fields for Normal Alkanes', *Journal of Physical Chemistry B* **102**(14), 2578–2586.

Chen, B., Potoff, J. J. and Siepmann, J. I. (2000), 'Adiabatic Nuclear and Electronic Sampling Monte Carlo Simulations in the Gibbs Ensemble: Application to Polarizable Force Fields for Water', *Journal of Physical Chemistry B* **104**(10), 2378–2390.

Chen, B., Potoff, J. J. and Siepmann, J. I. (2001), 'Monte Carlo Calculations for Alcohols and Their Mixtures with Alkanes. Transferable Potentials for Phase Equilibria. 5. United-Atom Description of Primary, Secondary, and Terniary Alcohols', *Journal of Physical Chemistry B* **105**(15), 3093–3104.

Chen, B. and Siepmann, J. I. (1999), 'Transferable Potentials for Phase Equilibria. 3. Explicit-Hydrogen Description of Normal Alkanes', *Journal of Physical Chemistry B* **103**(25), 5370–5379.

Chen, B. and Siepmann, J. I. (2000), 'A Novel Monte Carlo Algorithm for Simulating Strongly Associating Fluids: Applications to Water, Hydrogen Fluoride, and Acetic Acid', *Journal of Physical Chemistry B* **104**(36), 8725–8734.

Chen, B. and Siepmann, J. I. (2001), 'Improving the Efficiency of the Aggregation-Volume-Bias Monte Carlo Algorithm', *Journal of Physical Chemistry B* **105**(45), 11275–11282.

Coniglio, L. and Nouviaire, A. (2001), 'A Method for Estimating the Normal Boiling Point of Heavy Hydrocarbons Suitable for a Group-Contribution-Based Equation of State', *Industrial and Engineering Chemistry Research* **40**(7), 1781–1790.

Coniglio, L., Trassy, L. and Rauzy, E. (2000), 'Estimation of Thermophysical Properties of Heavy Hydrocarbons through a Group Contribution Based Equation of State', *Industrial and Engineering Chemistry Research* **39**(12), 5037–5048.

Crampon, C., Trassy, L., Avaullee, L., Neaua, E. and Coniglio, L. (2004), 'Simplification and extension of a predictive group contribution method for estimating heavy organic pure compound vapor pressures I. Hydrocarbons', *Fluid Phase Equilibria* **216**, 95–109.

DDB - Dortmund Data Bank (2004).

de Leeuw, S. W., Perram, J. W. and Smith, E. R. (1980*a*), 'Simulation of Electrostatic Systems in Periodic Boundary Conditions. I. Lattice Sums and Dielectric Constants', *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **373**(1752), 27–56.

de Leeuw, S. W., Perram, J. W. and Smith, E. R. (1980*b*), 'Simulation of Electrostatic Systems in Periodic Boundary Conditions. II. Equivalence of Boundry Conditions', *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **373**(1752), 57–66.

de Leeuw, S. W., Perram, J. W. and Smith, E. R. (1983), 'Simulation of Electrostatic Systems in Periodic Boundary Conditions. III. Further Theory and Applications ', *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **388**(1794), 173–193.

de Pablo, J. J., Laso, M. and Suter, U. W. (1992*a*), 'Estimation of the chemical potential of chain molecules by simulation', *Journal of Chemical Physics* **96**(8), 6197–6162.

de Pablo, J. J., Laso, M. and Suter, U. W. (1992*b*), 'Simulation of polyethylene above and below the melting point', *Journal of Chemical Physics* **96**(3), 2395–2403.

Democritus - Radial Distribution Function (2004). www.cse.clrc.ac.uk/msi/software/democritus/theory/rdf.html.

Dodd, L. R., Boone, T. D. and Theodorou, D. N. (1993), 'A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses', *Molecular Physics* **78**(4), 961–996.

Errington, J. R. and Panagiotopoulos, A. Z. (1998), 'Phase equilibria of the modified Buckingham exponential-6 potential from Hamiltonian scaling grand canonical Monte Carlo', *Journal of Chemical Physics* **109**(3), 1093–1100.

Errington, J. R. and Panagiotopoulos, A. Z. (1999*a*), 'A New Intermolecular Potential Model for the n-Alkane Homologous Series', *Journal of Physical Chemistry B* **103**(30), 6314–6322.

Errington, J. R. and Panagiotopoulos, A. Z. (1999*b*), 'New intermolecular potential models for benzene and cyclohexane', *Journal of Chemical Physics* **111**(21), 9731–9738.

Escobedo, F. A. (1998), 'Novel pseudoensembles for simulation of multicomponent phase equilibria', *Journal of Chemical Physics* **108**(21), 8761–8772.

Escobedo, F. A. (1999), 'Tracing coexistence lines in multicomponent fluid mixtures by molecular simulation', *Journal of Chemical Physics* **110**(24), 11999–12010.

Esselink, K., Loyens, L. D. J. C. and Smit, B. (1995), 'Parallel Monte Carlo simulations', *Physical Review E* **51**(2), 1560–1568.

Ferrenberg, A. M. and Swendsen, R. H. (1988), 'New Monte Carlo Technique for Studing Phase Transitions', *Physical Review Letters* **61**(23), 2635–2638.

Ferrenberg, A. M. and Swendsen, R. H. (1989), 'Optimized Monte Carlo Data Analysis', *Physical Review Letters* **63**(12), 1195–1198.

Fincham, D. (2000), Optimisation of the Ewald Sum. CCCP5.

Frenkel, D., Mooij, G. C. A. M. and Smit, B. (1991), 'Novel scheme to study structual and thermal properties of continuously deformable molecules', *Journal of Physics: Condensed Matter* **3**, 3053–3076.

Frenkel, D. and Smit, B. (2002), *Understanding Molecular Simulations: From Algorithms to Applications*, Vol. 1 of *Computational Science Series*, 2nd edn, Academic Press, San Diego.

Harris, R. A. (2004), Robust Equipment for the Measurement of Vapour-Liquid Equilibrium at High Temperatures and High Pressures, PhD thesis, University of KwaZulu-Natal.

Hummer, G. and Grønbech-Jensen, N. (1998), 'Pressure calculation in polar and charged systems using Ewald summation: Results for the extended simple point charge model of water', *Journal of Chemical Physics* **109**(7), 2791–2797.

James, F. (1990), 'A review of pseudorandom number generators', *Computer Physics Communications* **60**, 329–344.

James, F. (1994), 'RANDLUX: A Fortran implementation of the high-quality pseudorandom number generator of Lüscher', *Computer Physics Communications* **79**, 111–114.

Jorgensen, W. L. (1983), 'Theoretical Studies of Medium Effects on Conformational Equilibria', *Journal of Physical Chemistry* **87**(26), 5304–5314.

Jorgensen, W. L. (1986), 'Optimized Intermolecular Potential Functions for Liquid Alcohols', *Journal of Physical Chemistry* **90**(7), 1276–1284.

Jorgensen, W. L., Madura, J. D. and Swenson, C. J. (1984), 'Optimized Intermolecular Potential Functions for Liquid Hydrocarbons', *Journal of the American Chemical Society* **106**(22), 6638–6646.

KDB - Korea Thermophysical Properties Data Bank (2003). http://infosys.korea.ac.kr/kdb/.

Khare, R., Sum, A. K., Nath, S. K. and de Pablo, J. J. (2004), 'Simulation of the Vapor-Liquid Phase Equilibria of Primary Alcohols and Alcohol-Alkane Mixtures', *Journal of Physical Chemistry B* **108**(28), 10071–10076.

Kiyohara, K., Gubbins, K. E. and Panagiotopoulos, A. Z. (1997), 'Phase coexistence properties of polarizable Stockmayer fluids', *Journal of Chemical Physics* **106**(8), 3338–3347.

Kofke, D. A. (1993*a*), 'Direct evaluation of phase coexistence by molecular simulation via integration along the saturation line', *Journal of Chemical Physics* **98**(5), 4149–4162.

Kofke, D. A. (1993*b*), 'Gibbs-Duhem integration: a new method for direct evaluation of phase coexistence by molecular simulation', *Molecular Physics* **78**(6), 1331–1336.

Kofke, D. A. (2003), CE 530 Molecular Simulation - Course Notes. http://www.cheme.buffalo.edu/courses/ce530/index.html.

Loyens, L. D. J. C., Smit, B. and Esselink, K. (1995), 'Parallel Gibbs-ensemble simulations', *Molecular Physics* **86**(2), 171–183.

Lüscher, M. (1994), 'A portable high-quality random number generator for lattice field theory simulations', *Computer Physics Communications* **79**, 100–110.

Martin, M. G. (2004*a*), MCCCS TOWHEE . http://towhee.sourceforge.net.

Martin, M. G. (2004*b*), Personal Communication.

Martin, M. G. and Siepmann, J. I. (1998), 'Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes', *Journal of Physical Chemistry B* **102**(14), 2569–2577.

Martin, M. G. and Siepmann, J. I. (1999), 'Novel Configuration-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description of Branched Alkanes', *Journal of Physical Chemistry B* **103**(21), 4508–4517.

McDonald, I. R. (1972), 'NpT-ensemble Monte Carlo calculations for binary liquid mixtures', *Molecular Physics* **23**(1), 41–58.

McGlashan, M. L. and Potter, D. J. B. (1962), 'An Apparatus for the Measurement of the Second Virial Coefficients of Vapours; The Virial Coefficients of Some n-alkanes and Some Mixtures of n-alkanes.', *Proceeding of the Royal Society of London. Series A, Mathematical and Physical Sciences* **267**(1331), 478–500.

McKnight, T. (2003), Personal Communication.

McKnight, T. (2005), Molecular Simulation of Binary VLE using Beowulf Clusters, PhD thesis, University of KwaZulu-Natal.

McKnight, T., Ramjugernath, D. and Starzak, M. (2002), Simulating Vapour-Liquid Phase Equilibria using NERD, TraPPE-UA and Buckingham Exp-6 United-Atom Potentials.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M., Teller, A. H. and Teller, E. (1953), 'Equation of state calculations by fast computing machines', *Journal of Chemical Physics* **21**, 1087–1092.

Mooij, G. C. A. M. and Frenkel, D. (1996), 'A Systematic Optimization Scheme for Configurational Bias Monte Carlo', *Molecular Simulation* **17**, 41–55.

Nath, S. K. (2003), 'Molecular Simulation of Vapor-Liquid Phase Equilibria of Hydrogen Sulphide and Its Mixtures with Alkanes', *Journal of Physical Chemistry B* **107**(35), 9498–9504.

Nath, S. K., Banaszak, B. J. and de Pablo, J. J. (2001*a*), 'A new united atom force field for α-olefins', *Journal of Chemical Physics* **114**(8), 3612–3616.

Nath, S. K., Banaszak, B. J. and de Pablo, J. J. (2001*b*), 'Simulation of Ternary Mixtures of Ethylene, 1-Hexene, and Polyethylene', *Macromolecules* **34**(22), 7841–7848.

Nath, S. K. and de Pablo, J. J. (2000), 'Simulation of vapour-liquid equilibria for branched alkanes', *Molecular Physics* **98**(4), 231–238.

Nath, S. K., Escobedo, F. A. and de Pablo, J. J. (1998), 'On the simulation vapor-liquid equilibria for alkanes', *Journal of Chemical Physics* **108**(23), 9905–9911.

Nath, S. K., Escobedo, F. A., de Pablo, J. J. and Patramai, I. (1998), 'Simulation of Vapor-Liquid Equilibria for Alkane Mixtures', *Industrial and Engineering Chemistry Research* **37**(8), 3195–3202.

Nath, S. K. and Khare, R. (2001), 'New forcefield parameters for branched hydrocarbons', *Journal of Chemical Physics* **115**(23), 10837–10844.

National Supercomputer Center, Linköping University, Sweden (2003). http://www.nsc.liu.se/.

NIST Chemistry WebBook (2003). http://webbook.nist.gov/chemistry/.

Nosé, S. and Klein, M. L. (1983), 'Constant pressure molecular dynamics for molecular systems', *Molecular Physics* **50**, 1055–1076.

Page, C. G. (1988), *Professional Programmer's Guide to Fortran77*, Pitman. Text made freely available online by author. Copyright 1988, 2001 Clive G. Page http://www.star.le.ac.uk/~cgp.

Panagioropoulos, A. Z., Wong, V. and Floriano, M. A. (1998), 'Phase Equilibria of Lattice Polymers from Histogram Reweighting Monte Carlo Simulations', *Macromolecules* **31**(3), 912–918.

Panagiotopoulos, A. Z. (1987), 'Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble', *Molecular Physics* **61**(4), 813–826.

Panagiotopoulos, A. Z. (2000), 'Monte Carlo methods for phase equilibria of fluids', *Journal of Physics: Condensed Matter* **12**, R25–R52.

Panagiotopoulos, A. Z., Quirke, N., Stapleton, M. and Tildesley, D. J. (1988), 'Phase equilibria by simulation in the Gibbs ensemble. Alternative derivation, generalization and application to mixture and membrane equilibria', *Molecular Physics* **63**(4), 527–545.

Potoff, J. J. and Panagiotopoulos, A. Z. (1998), 'Critical point and phase behaviour of the pure fluid and a Lennard-Jones mixture', *Journal of Chemical Physics* **109**(24), 10914–10920.

Rao, Y. V. C. (1994), *Postulational and Statistical Thermodynamics*, Allied Publishers Limited.

Ray, J. R. (1991), 'Microcanonical ensemble Monte Carlo method', *Physical Review A* **44**(6), 4061–4064.

Rosenbluth, M. N. and Rosenbluth, A. W. (1955), 'Monte Carlo Calculation of the Average Extension of Molecular Chains', *Journal of Chemical Physics* **23**(2), 356–359.

Sadus, R. J. (1999), *Molecular Simualtion of Fluids: Theory, Algorithms and Object-Orientation*, Elsevier.

Shi, W. and Johnson, J. K. (2001), 'Histogram reweighting and finite-size scaling study of the Lennard-Jones fluids', *Fluid Phase Equilibria* **187–188**, 171–191.

Siepmann, J. I. (1990), 'A Method for the direct calculation of chemical potentials for dense chain systems', *Molecular Physics* **70**(6), 1145–1158.

Siepmann, J. I. and Frenkel, D. (1992), 'Configurational bias Monte Carlo: a sampling scheme for flexible chains', *Molecular Physics* **75**(1), 59–70.

Siepmann, J. I., Karaborni, S. and Smit, B. (1993), 'Simulating the critical behaviour of complex fluids', *Nature* **365**, 330–332.

Smit, B., Karaborni, S. and Siepmann, J. I. (1995), 'Computer simulations of vapor-liquid phase equilibria of n-alkanes', *Journal of Chemical Physics* **102**(5), 2126–2140.

Smit, B., Karaborni, S. and Siepmann, J. I. (1998), 'Erratum: "Computer simulations of vapor-liquid phase equilibria of n-alkanes" [J. Chem. Phys. 102, 2126 (1995)]', *Journal of Chemical Physics* **109**(1), 352.

Smith, B. D. and Srivastava, R. (1986*a*), *Thermodynamic Data for Pure Compounds: Part A Hydrocarbons, and Ketones*, Elsevier, Amsterdam.

Smith, B. D. and Srivastava, R. (1986*b*), *Thermodynamic Data for Pure Compounds: Part B Halogenated Hydrocarbons and Alcohols*, Elsevier, Amsterdam.

Spencer, C. F. and Danner, R. P. (1972), 'Improved Equation for Prediction of Saturated Liquid Density', *Journal of Chemical and Engineering Data* **17**(2), 236–241.

Toxvaerd, S. (1990), 'Molecular dynamics calculation of the equation of state of alkanes', *Journal of Chemical Physics* **93**(6), 4290–4295.

Ungerer, P. (2003), 'From Organic Geochemistry to Statistical Thermodynamics: the Development of Simulation Methods for the Petroleum Industry', *Oil and Gas Science and Technology – Rev. IFP* **58**(2), 271–297.

Ungerer, P., Beauvais, C., Delhommelle, J., Boutin, A., Rousseau, B. and Fuchs, A. H. (2000), 'Optimization of the anisotropic united atoms intermolecular potential for n-alkanes', *Journal of Chemical Physics* **112**(12), 5499–5510.

Valleau, J. P. and Cohen, L. K. (1980), 'Primitive model electrolytes. I. grand canonical Monte Carlo computations', *Journal of Chemical Physics* **72**, 5935–5941.

van den Berg, A. W. C., Bromley, S. T., Ramsahye, N. and Maschmeyer, T. (2004), 'Diffusion of Molecular Hydrogen through Porous Materials: The Importance of Framework Flexibility', *Journal of Physical Chemistry B* **108**(16), 5088–5094.

Vargaftik, N. B. (1975), *Tables on the Thermophysical Properties of Liquids and Gases - In Normal and Dissociated States*, 2nd edn, Hemisphere Publishing Corporation, Washington.

Vlugt, T. J. H. (2000), Adsorption and Diffusion in Zeolites: A Computational Study, PhD thesis, University of Amsterdam.

Vlugt, T. J. H., Krishna, R. and Smit, B. (1999), 'Molecular Simulations of the Adsorption Isotherms for Linear and Branched Alkanes and Their Mixtures in Silicate', *Journal of Physical Chemistry B* **103**(7), 1102–1118.

Vlugt, T. J. H., Martin, M. G., Smit, B., Siepmann, J. I. and Krishna, R. (1998), 'Improving the efficiency of the configurational-bias Monte Carlo algorithm', *Molecular Physics* **94**(4), 727–733.

Vlugt, T. J. H., Zhu, W., Kapteijn, F., Moulijn, J. A., Smit, B. and Krishna, R. (1998), 'Adsorption of Linear and Branched Alkanes in the Zeolite Silicalite-1', *Journal of American Chemical Society* **120**(22), 5599–5600.

Wick, C. D., Martin, M. G. and Siepmann, J. I. (2000), 'Transferable Potentials for Phase Equilibria. 4. United-Atom Description of Linear and Branched Alkenes and Alkylbenzenes', *Journal of Physical Chemistry B* **104**(33), 8008–8016.

Widom, B. (1963), 'Some topics in the theory of fluids', *Journal of Chemical Physics* **39**, 2808–2812.

Widom, B. (1982), 'Potential-Distribution Theory and the Statistical Mechanics of Fluids', *Journal of Physical Chemistry* **86**(6), 869–872.