MOLECULAR SIMULATION AND MODELING OF THE PHASE EQUILIBRIA OF POLAR COMPOUNDS

by

Scott Llewellyn Clifford

[MSc.(Eng.)]

University of KwaZulu-Natal, Durban

Submitted in fulfillment of the academic requirements for the degree "Doctor of Philosophy in Engineering" at the School of Chemical Engineering, University of KwaZulu-Natal

Durban

2006

Preface

The work presented in this thesis entitled "Molecular Simulation and Modeling of the Phase Equilibria of Polar Compounds" was performed at the School of Chemical Engineering, University of KwaZulu-Natal, Howard College Campus, from February 2004 to June 2006. The project was supervised by Prof. D. Ramjugernath (University of KwaZulu-Natal) and Dr. K. Bolton (University College of Borås).

This thesis is submitted as the full requirement for the degree "Doctor of Philosophy in Chemical Engineering". To the best of my knowledge, all of the work in this thesis is original, except where due reference is made in the text. It has not (in whole or in part) been previously submitted to any tertiary institution as part of a degree.

Scott Llewellyn Clifford
23 June 2006

Acknowledgements

I would like to take this opportunity to acknowledge and thank the many people and organizations who have made a significant contribution to this project:

- My supervisors, Prof. Deresh Ramjugernath (University of KwaZulu-Natal) and Dr. Kim Bolton (University College of Borås) for their ideas, support and guidance over the past two and a half years of research.

- Prof. J. David Raal for his invaluable assistance with the modeling (Chemical Theory) section of the thesis.

- Tyrone McKnight and Nicholas du Preez, whose help and advice during the course of this project were much appreciated.

- Anders Askåsen, Daniel Johansen and Peter Ahlström (University College of Borås) for their help and friendship.

- The Swedish National Allocations Committee for granting access to the Swedish national supercomputing facilities (Monolith) in Linköping.

- The financial support of the Swedish International Co-operation Development Agency (SIDA), SASOL, THRIP, the South African National Research Foundation Thuthuka Programme and the South African National Research Foundation International Science Liaison.

- My friends and colleagues from the School of Chemical Engineering, in particular: Tyrone, Nick, Roger, John-Roy, Jason, Etienne, Vivek, Allen, Yash and Minal.

- My family, for their love and support during my many years of study.

- God, without Whom none of this would be possible.

Abstract

The initial phase of the project involved an investigation into the modeling of binary carboxylic acid vapour-liquid equilibrium (VLE) data. This stemmed from the Masters research that led into the current study, in which the conventional gamma-phi formulation of VLE was found to inadequately describe the complicated acid chemistry. In an effort to correctly describe the dimerization occurring in both the liquid and vapour phases, the chemical theory of vapour-phase imperfections was applied. The chemical theory technique allowed the experimental liquid-phase activity coefficients to be accurately calculated by taking the vapour phase dimerization into account. Once these activity coefficients had been determined, standard Gibbs excess energy models were fitted to permit analysis of the VLE data's thermodynamic consistency. In addition, the typical bubble-point iteration scheme used for VLE data regression was adapted to include the chemical theory expressions necessary for satisfactory modeling of the carboxylic acids.

The primary focus of this study was to determine the ability of currently available computer simulation techniques and technology to correctly predict the phase equilibria of polar molecules. Thus, Monte Carlo simulations in the *NVT*- and *NPT*-Gibbs ensembles were used to predict pure component and binary phase equilibrium data (respectively), for a variety of polar compounds. The average standard deviations for these simulation results lay between 1 and 2 % for the saturated liquid densities, and varied between 5 and 10 % for the saturated vapour pressures and densities.

Pure component data were simulated for alcohols, carboxylic acids, hydrogen sulfide (H_2S), sulfur dioxide (SO_2) and nitrogen dioxide (NO_2). For H_2S, SO_2 and NO_2, a potential model parameterized as part of this project was used to describe the molecular interactions. All the other compounds were simulated using the TraPPE-UA force field. The simulation results for the alcohols and acids showed a consistent saturated vapour pressure over-prediction of 5 - 20 % depending on the species and the system temperature. The liquid density predictions were, in general, good and on average differed from experiment by 1 - 2 %. The critical temperatures and densities were estimated from the pure component data by fitting to the scaling law and the law of rectilinear diameters. They were found to lie within 1 and 2 % of the experimental

values for the carboxylic acids and alcohols, respectively. Clausius-Clapeyron plots of the saturated vapour pressures allowed the critical pressure and normal boiling points to be determined. The critical pressures were, as expected, over-predicted for both compound classes and the normal boiling points were under-estimated somewhat for the acids, but deviated from experiment by less than 0.5 % for the alcohols.

A Lennard-Jones 12-6 plus Coulombic potential energy surface was parameterized for H_2S, SO_2 and NO_2. For H_2S, the proposed force field offers improved saturated vapour pressure and vapour density predictions when compared to the existing NERD force field, and comparable accuracy with the recent models of Kamath and co-workers. SO_2 and NO_2 had not previously been parameterized for a Lennard-Jones 12-6 based force field. For SO_2, there was excellent agreement with experimental data. In the case of NO_2, the saturated liquid density predictions were very good, but the vapour pressures and densities were over-predicted.

Binary VLE simulations were carried out for systems consisting purely of carboxylic acids, and also for H_2S and SO_2 with a selection of alkanes and alcohols. The liquid and vapour composition predictions were good for the acid systems, but the anticipated pressure and temperature deviations were observed in the isothermal and isobaric simulations, respectively. The H_2S + alkane systems were generally good, as were the SO_2 + alkane systems. For both H_2S and SO_2, the systems involving an alcohol displayed a characteristic pressure over-estimation. The azeotropes were, in most cases, predicted fairly well; the exception was the SO_2 + methane binary.

A sensitivity analysis of the Lennard-Jones unlike interaction parameters was also conducted. It was demonstrated that even minor changes to these parameters can have a significant effect on the final simulation results. The considerable affect that these parameters have on the simulation outputs was emphasized by studying the influence of different combining rules on the H_2S + methane and H_2S + ethane binary systems.

Analysis of the radial distribution functions indicated that hydrogen bonding and dimerization were occurring in the alcohol and carboxylic acid systems, respectively. The H_2S, SO_2 and NO_2 distribution functions showed little sign of any association, except for a small plateau in that of SO_2. A radial distribution function from one of the carboxylic acid binary simulations was also analysed, and supported the assumption made in the chemical theory modeling work of using a geometric mean (instead of twice the geometric mean, which is favoured by some researchers) to determine the heterodimerization constant, K_{AB}.

Table of Contents

List of Figures

List of Tables

Nomenclature

The project was divided into two main sections, one dealing with chemical theory modeling of carboxylic acids and the other with molecular simulation of polar compounds. A number of symbols from the different areas of study overlapped, and, hence, the nomenclature is also divided into two sections.

Chemical Theory Symbols

English Letters

A	Component 1 in a binary carboxylic acid mixture
A_K	Constant in the chemical equilibrium constant equation
A_1, A_2	Monomer and dimer of component 1
A'	Antoine vapour pressure equation constant
a	Constant in the temperature-dependence equation for λ_{ij}
a_{i_1}	Activity of the true monomer species
AB	Heterodimer in a binary carboxylic acid mixture
B	Component 2 in a binary carboxylic acid mixture
B_K	Constant in the chemical equilibrium constant equation
B_1, B_2	Monomer and dimer of component 2
B'	Antoine vapour pressure equation constant
b	Constant in the temperature-dependence equation for λ_{ij}
C'	Antoine vapour pressure equation constant
c	Constant in the temperature-dependence equation for λ_{ij}
H^E	Excess enthalpy
K	Chemical equilibrium constant
K_{ii}	Vapour-phase chemical equilibrium constant for component i
K_{AB}	Vapour-phase chemical equilibrium constant for heterodimer
n_1, n_2	Number of moles of components 1 and 2
n_i	Number of moles of species i

P	System pressure
P_c	Critical pressure
P_A^{sat}, P_B^{sat}	Saturation pressures for the pure carboxylic acid species, A and B
$P_{A_1}^{sat}$, $P_{B_1}^{sat}$	Saturation pressures for the carboxylic acid monomers, A_1 and B_1
R	Universal gas constant
T	Temperature (K)
t	Temperature (°C)
T_c	Critical temperature
V	Volume
V_j, V_k	Molar volumes of pure liquids j and k, respectively
V_{jk}	Molar volume ratio
V_c	Critical volume
V^E	Excess volume
x_1, x_2	Liquid-phase mole fractions of components 1 and 2 (for the original TK-Wilson equation; Equation 2.23)
x_A, x_B	Analytical liquid-phase mole fractions of components A and B
y_A, y_B	Analytical vapour-phase mole fractions of components A and B
Z_c	Critical compressibility factor

Greek Letters

γ	Liquid-phase activity coefficient
γ_A, γ_B	Activity coefficients for the pure species A and B
γ_{A_1}, γ_{B_1}	Activity coefficients for the monomer species A_1 and B_1
δ	Residual (e.g. δP, δy)
ε	Tolerance in bubble point iteration scheme
ε_γ	Tolerance in the liquid-phase activity coefficient regression
η_i	Mole fraction of species i in the vapour phase
λ_{ij}, Λ_{ij}	Parameters in the T-K Wilson equation
ξ_i	Mole fraction of species i in the liquid phase
Φ_i	Vapour-phase correction factor for component i

Molecular Simulation Symbols

English Letters

A	Thermodynamic property of interest
b	Bond length
b_0	Equilibrium bond length
b_i	i^{th} trial orientation
c_i	Constants in the TraPPE-UA torsional potential (where i = 0, 1, 2, 3)
C	Normalization constant in the Gibbs ensemble CBMC algorithm
d	Distance from a molecule's center of mass to an interaction site
E	Total energy
E_H	Hermholtz free energy
E_G	Gibbs free energy
f_1	Angle in torsional potential
F	Degrees of freedom in a thermodynamic system
$g(r)$	Pair distribution function (in this work, this function is radial in nature)
h	Planck's constant (6.626×10^{-34} J.s)
$H(\mathbf{p},\mathbf{q})$	Hamiltonian of the thermodynamic system
i	Atom (or site or segment) i
I	Integral
j	Atom (or site or segment) j
k	Number of trial segments
k_b	Harmonic constant for bond stretching
k_B	Stefan Boltzmann constant (5.6705×10^{-8} W/m^2.K^4)
k_θ	Harmonic constant for angle bending
K	Kinetic energy
K_{max}	Maximum number of inverse space vectors
l	Number of regrown molecular segments
L	Simulation box length
L^*	Molecule aspect ratio
m	Molecule chain
M_i	Molecule i
n	New system configuration

N	Number of particles (or molecules)
o	Old system configuration
p	Particle momentum
P	Pressure
P_v	Virial pressure
P_T	Thermodynamic pressure
q	Particle coordinates
q_i	Atomic charge of site i
Q	Partition function
r_{cut}	Cutoff radius
r_{cut}^*	Inner cutoff radius
r_{ij}	Distance between molecular centers (molecular separation)
$\mathbf{s}$	Scaled coordinate vector
S	Entropy
t	Time
T	Temperature
$u(x)$	Nonnegative, nondecreasing function of x
U	Potential energy
V	Volume
V_i	Constants in the OPLS-UA torsional potential (where i = 1, 2, 3)
V_o	Basic volume unit in isothermal-isobaric ensemble
w	Rosenbluth weight
$w(r_{ij})$	Intermolecular pair virial function
W	Rosenbluth factor
W_v	Internal virial
x	Liquid phase composition
y	Vapour phase composition
Z	Number of atom types in a system

Greek Letters

α	Parameter determining the Gaussian distribution width
$\alpha(o \rightarrow n)$	Transition matrix (contains the probabilities of a certain move being performed between state o and state n)

β	Reciprocal temperature $(1/k_B.T)$
β_s	Universal Ising exponent, critical exponent or scaling exponent
β_{eff}	Effective scaling exponent
ε	Pair potential energy parameter
ε_s	Relative permittivity
κ	Square root of α
σ	Pair potential size parameter
θ	Bond angle
θ_0	Equilibrium bond angle
ϕ	Dihedral angle
τ	Total experiment time
$\pi(o \rightarrow n)$	Probability of moving from state o to state n
$\Xi(o \rightarrow n)$	Probability of a trial move from state o to state n being accepted
μ	Chemical potential
λ	Wavelength
Λ	Thermal de Broglie wavelength
ρ	Density
ρ_n	Average number density
ρ_{pf}	Density probability function
ζ	Total number of uniformly distributed random values taken from $u(x)$
P	Probability
χ_i	Gaussian screen charge at site i (Ewald sum)

Subscript Symbols

A	Angle bending
b	Boiling point
bulk	Bulk (or mean) property
c	Critical point
ens	Ensemble
gen	Generation
Gibbs	Gibbs ensemble
i	i^{th} particle (or atom or site or segment)

int	Internal
I	Phase 1
II	Phase 2
j	j^{th} particle (or atom or site or segment)
K	Kinetic
l	Liquid phase
L	Lennard-Jones
n	Refers to state *n*
o	Refers to state *o*
obs	Observed
pf	Probability function
swap	Swap move
test	Test particle (or molecule)
torsion	Torsional potential
T	Torsional
v	Vapour phase

Superscript Symbols

ang	Angle bending
bond	Bonded interactions
charge	Charged interactions
ex	Excess
ext	External
HS	Hard-sphere
int	Internal
LJ	Lennard-Jones
LJ-shift	Shifted Lennard-Jones potential
nonbond	Non-bonded interactions
pos	Position
or	Orientation
SS	Soft-sphere
SW	Square-well
tail	Tail correction

tors	Torsional
tot	Total
+	Test (or ghost) particle

Other Symbols

<...>	Average property
COM	Center of mass
erfc	Complementary error function

It would appear that we have reached the limits of what it is possible to achieve with computer technology, although one should be careful with such statements, as they tend to sound pretty silly in five years.

John von Neumann (ca. 1949)

I think there's a world market for about five computers.

Thomas Watson (founder of IBM)

Statistics: the only science that enables different experts using the same figures to draw different conclusions.

Evan Esar

1

Introduction

The science of thermodynamics is a well-established discipline that has existed since the early nineteenth century. Originally, thermodynamics was developed in an effort to describe the operation and capabilities of steam engines (Smith et al., 1996). This origin is evident in the name, which (loosely translated) means "movement from heat". Scientists soon discovered, however, that the laws of thermodynamics derived for simple steam engines pertained to many scientific fields. Indeed, thermodynamics currently finds application in most facets of modern engineering.

This is particularly true in the area of separations technology (distillation, extraction, adsorption etc.). A tremendous amount of research and effort has been poured into the development of thermodynamic relationships and techniques that improve the design, operation and maintenance of separation processes and equipment. To effectively improve and develop the thermodynamic theory pertinent to separations technology, accurate, reliable phase equilibrium data are required. In the past, the only means of obtaining such data was through laboratory measurement and experimentation. The advent of computers, however, has introduced alternative methods (such as molecular simulation) for generating phase equilibrium data.

Molecular simulation is a relatively new field of research, having existed in its current form since the early 1980's. The exponential growth of computing power over the last twenty years has catapulted computer modeling and simulation to the

forefront of scientific endeavour. Combined with the simultaneous improvement in the accuracy and complexity of the potential models used to describe interactions in molecular simulation, this has brought us to the point where computer simulations are now capable of rivaling traditional experimental methods for acquiring experimental data. However, the data produced must achieve a certain level of accuracy and reliability before it can be confidently used in the design and development of industrial processes. For molecular simulation, the requisite precision is still lacking for complex molecules. Polar compounds are particularly challenging because they tend to agglomerate into clusters. Furthermore, they are often capable of forming hydrogen bonds, which lead to phenomena such as dimers, polymers and heterodimers.

The principle aim of this project, therefore, was to determine the ability of contemporary molecular simulation methods to accurately predict the phase equilibrium of polar molecules. To this end, pure component phase equilibrium data were simulated for alcohols, carboxylic acids, hydrogen sulfide, sulfur dioxide and nitrogen dioxide, and binary vapour-liquid equilibrium (VLE) data were simulated for systems involving (primarily) carboxylic acids, hydrogen sulfide and sulfur dioxide. The hydrogen sulfide, sulfur dioxide and nitrogen dioxide simulations were carried out using a Lennard-Jones based potential energy model, parameterized as part of this project. During this parameterization process, the sensitivity of the Lennard-Jones parameters to minor perturbations was studied.

Carboxylic acids are a class of polar compound that hold considerable economic importance. However, due to their complicated and difficult chemistry, they have received little attention from researchers. During the Masters project that preceded the current study, several carboxylic acid binary VLE systems were measured and modeled using the conventional gamma-phi formulation of VLE. The liquid-phase activity coefficient models (such as the UNIQUAC, NRTL and Wilson equations) were found to provide an inadequate description of the carboxylic acid VLE due to the extensive dimerization that occurs in both the vapour and liquid phases. To correctly account for the effect of the dimers (and, in some cases, trimers), the chemical theory of vapour phase imperfections was used in this work to determine the experimental liquid-phase activity coefficients. Thus, the secondary focus of the project was the accurate modeling of the polar carboxylic acids.

The thesis is arranged in the following manner: in Chapter 2, the chemical theory of vapour phase imperfections is introduced and derived for the carboxylic acid systems.

A novel bubble-point algorithm is developed and presented that incorporates the chemical theory necessary for satisfactory modeling of the carboxylic acid VLE. Chapter 3 deals with the statistical mechanics theory that underpins the Monte Carlo simulation method, which is described in detail in Chapter 4. In Chapter 5, the molecular interactions used to determine the potential energy during a simulation are discussed. The Gibbs ensemble technique that is exclusively used through the simulation portion of this project is described in Chapter 6. Chapter 7 introduces the computer hardware and software required for the successful implementation, running and maintenance of a Beowulf cluster (used to run the computer simulations). In addition, a brief discussion of some common problems (and, where possible, the corresponding solutions) encountered during the course of this study is included. The results and analysis of the work conducted during this project are presented in the discussion chapter, Chapter 8. Chapters 9 and 10 contain, respectively, the conclusion and the recommendations arising from this work.

2

Chemical Theory

Polar compounds capable of forming strongly hydrogen-bonded complexes such as dimers, polymers and heterodimers introduce considerable complexity in the analysis and interpretation of vapour-liquid equilibrium (VLE) data. This is particularly true of carboxylic acids, for which extensive dimerization and cross-dimerization occurs in both the liquid and vapour phases, even at very low pressures (Prausnitz et al., 1980).

Consequently, the liquid-phase activity coefficients cannot be determined from the commonly used gamma-phi formulation of VLE. To account for the dimers that are present (in addition to the pure component monomers), a theory incorporating these complexes should be used instead. In addition, a satisfactory description of the highly non-ideal vapour phase is of particular importance if realistic liquid-phase activity coefficients are to be calculated from the measured P-T-x-y data. Thus, in this work the experimental liquid-phase activity coefficients were determined using a procedure based on the chemical theory for the vapour phase similar to those of Tamir and Wisniak (1975), Marek (1955), and Marek and Standart (1954).

Once the experimental liquid-phase activity coefficients were calculated, various common excess Gibbs energy models were fitted to the activity coefficients and the T-K Wilson model was found to be the most accurate. This permitted the thermodynamic consistency of the data to be tested. Furthermore, the bubble-point calculation procedure generally used in conjunction with the Gibbs excess models for modeling of

binary VLE data is adapted here to allow integration of the chemical theory necessary for successful modeling of carboxylic acid systems.

The content of this chapter is based to a large extent on Clifford et al. (2005).

2.1 Associating System Analysis

For a mixture of two carboxylic acids A and B, we assume that the following reversible dimerization reactions occur in the vapour phase at low to moderate pressures:

$$2A_1 \xleftrightarrow{K_{AA}} A_2 \tag{2.1}$$

$$2B_1 \xleftrightarrow{K_{BB}} B_2 \tag{2.2}$$

$$A_1 + B_1 \xleftrightarrow{K_{AB}} AB \tag{2.3}$$

The corresponding chemical equilibrium constants (that depend only upon temperature) are given by:

$$K_{AA} = \frac{\eta_{A_2}}{P\eta_{A_1}^2} \tag{2.4a}$$

$$K_{BB} = \frac{\eta_{B_2}}{P\eta_{B_1}^2} \tag{2.4b}$$

$$K_{AB} = \frac{\eta_{AB}}{P\eta_{A_1}\eta_{B_1}} \tag{2.4c}$$

The η_k (k = A, B, A_2, B_2, AB) are the true species mole fractions in the vapour phase and must be determined by material balance and equilibrium relations, as shown below. If the vapour phase is constituted from n_1 and n_2 stoichiometric moles of the acids A and B respectively, the following equations are readily derived:

$$\eta_{A_1} = \frac{n_1 - 2n_{A_2} - n_{AB}}{n_1 + n_2 - n_{A_2} - n_{AB} - n_{B_2}} \tag{2.5}$$

$$\eta_{B_1} = \frac{n_2 + 2n_{B_2} + n_{AB}}{n_1 + n_2 - n_{A_2} - n_{AB} - n_{B_2}} \tag{2.6}$$

$$\eta_{A_2} = \frac{n_{A_2}}{n_1 + n_2 - n_{A_2} - n_{AB} - n_{B_2}} \tag{2.7}$$

$$\eta_{B_2} = \frac{n_{B_2}}{n_1 + n_2 - n_{A_2} - n_{AB} - n_{B_2}} \tag{2.8}$$

$$\eta_{AB} = \frac{n_{AB}}{n_1 + n_2 - n_{A_2} - n_{AB} - n_{B_2}} \tag{2.9}$$

Introduction of Equations 2.4 into Equations 2.5 to 2.9 and normalization of the unknown mole numbers n_{A_2}, n_{B_2} and n_{AB} with respect to the total number of moles (n_1 + n_2), produces three equations in terms of the three mole number ratios and the stoichiometric (i.e. measured) vapour mole fractions y_A and y_B. These may, in principle, be solved by a suitable numerical procedure, subject to the constraint that the sum of the true mole fractions must add up to one and provided that the equilibrium constants, K, are known.

A more attractive and elegant method, however, is that proposed by Tamir and Wisniak (1975). It is readily established that the stoichiometric vapour-phase mole fractions are given by:

$$y_A = \frac{\eta_{A_1} + 2\eta_{A_2} + \eta_{AB}}{1 + \eta_{A_2} + \eta_{B_2} + \eta_{AB}} \tag{2.10}$$

$$y_B = \frac{\eta_{B_1} + 2\eta_{B_2} + \eta_{AB}}{1 + \eta_{A_2} + \eta_{B_2} + \eta_{AB}} \tag{2.11}^{\dagger}$$

---

† There is a typographical error in Equation 2.11 as given by Tamir and Wisniak (1975): they omitted the η_{B_2} term in the denominator.

Substitution of Equations 2.4 into Equations 2.10 and 2.11 yields two quadratic equations in the unknown true species monomer concentrations, η_{A_1} and η_{B_1} (Tamir and Wisniak, 1975):

$$K_{AA}P(1+y_B)\eta_{A_1}^2 + \eta_{A_1} - y_A\left(1+K_{BB}P\eta_{B_1}^2\right) + K_{AB}P\eta_{A_1}\eta_{B_1}y_B = 0 \qquad (2.12)$$

$$K_{BB}P(1+y_A)\eta_{B_1}^2 + \eta_{B_1} - y_B\left(1+K_{AA}P\eta_{A_1}^2\right) + K_{AB}P\eta_{A_1}\eta_{B_1}y_A = 0 \qquad (2.13)$$

Since the simultaneous solution of these equations produces multiple roots, their behaviour was studied for the measured systems. As shown in Figure 2.1, there is only a single intersection of the two equations in the first (positive) quadrant (at a particular y_A) and, hence, only one meaningful solution. These intersections yield the experimental values for η_{A_1} and η_{B_1} at the respective vapour-phase composition. Values for the remaining true species compositions, η_{A_2}, η_{B_2} and η_{AB} then follow directly from Equations 2.4.

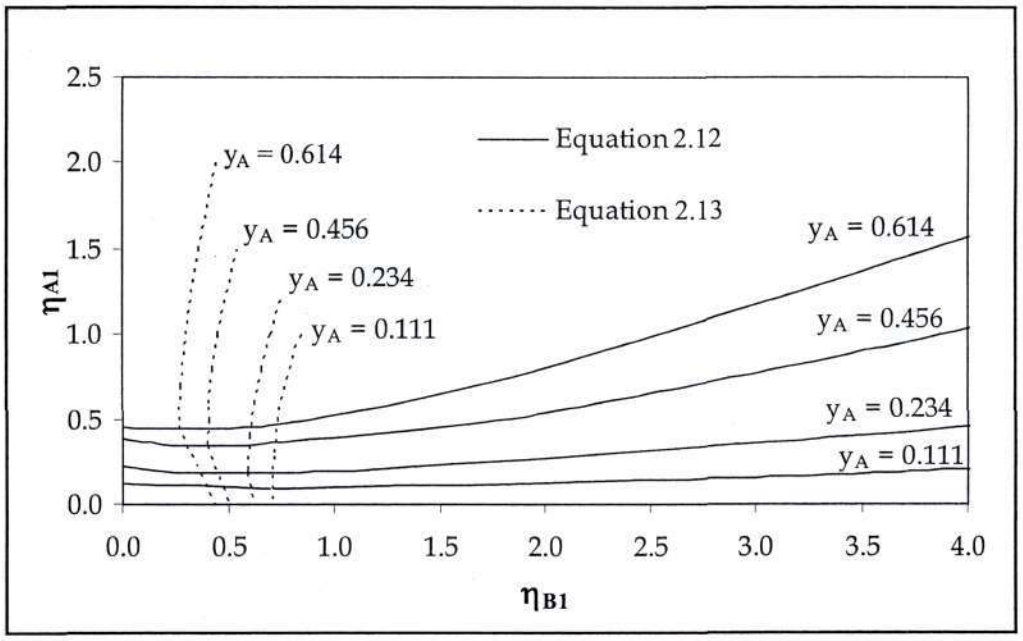

Figure 2.1: Graphical representation of Equations 2.12 and 2.13 showing that only a single, meaningful simultaneous solution exists at each vapour-phase composition, y_A.

With η_{A_1} and η_{B_1} determined for the full concentration range (and assuming ideal gas behaviour of the true species in the vapour phase) the equilibrium condition for monomer i is as follows (Tamir and Wisniak, 1975):

$$P\eta_{i_1} = \gamma_{i_1}\xi_{i_1}P_{i_1}^{sat} \tag{2.14}$$

where γ_{i_1} is the monomer activity coefficient in the liquid phase and ξ_{i_1} is the monomer mole fraction in the liquid phase. The vapour pressure of the monomer, $P_{i_1}^{sat}$, although not directly measurable, is readily found from the pure acid vapour pressure, P_i^{sat}, since K_{ii} is known (Tamir and Wisniak, 1975):

$$P_i^{sat} = P_{i_1}^{sat} + P_{i_2}^{sat} = P_{i_1}^{sat} + K_{ii}\left(P_{i_1}^{sat}\right)^2 \tag{2.15}$$

In Equation 2.15, P_i^{sat} is calculated from the Antoine equation:

$$\ln P_i^{sat} = A' - \frac{B'}{t + C'} \tag{2.16}$$

where P_i^{sat} is in kPa, t in °C and the Antoine parameters (Clifford et al., 2004) are given in Table A-1, Appendix A.

Solving the quadratic Equation 2.15 for the monomer vapour pressure, we have:

$$P_{i_1}^{sat} = \frac{-1 + \sqrt{1 + 4K_{ii}P_i^{sat}}}{2K_{ii}} \tag{2.17}$$

The derived experimental activity coefficients for the liquid phase, γ_i, can be obtained at a particular liquid-phase composition (x_i) from:

$$\gamma_i = \frac{a_{i_1}}{x_i} = \frac{\gamma_{i_1}\xi_{i_1}}{x_i} \tag{2.18}$$

Equation 2.18 is exact if the true species are in chemical equilibrium (Prausnitz et al., 1999) and states that the "apparent" activity equals the "true" activity of the monomer species. Re-arranging and combining Equation 2.18 with Equation 2.14 yields, finally:

$$P\eta_{i_1} = x_i\gamma_i P_{i_1}^{sat} \tag{2.19}$$

Once the true species concentrations have been found via simultaneous solution of Equations 2.12 and 2.13, the experimental liquid-phase activity coefficients are readily obtained using Equation 2.19.

An attractive feature of the Tamir and Wisniak (1975) approach is that values for the true species liquid-phase activity coefficients, γ_{i_1}, need not be calculated or assumed. Since the liquid-phase equilibrium constants are not generally known and are not easy to determine (see below), it is difficult (if not impossible) to find the true monomer species activity coefficient γ_{i_1} independently.

2.2 Chemical Equilibrium Constants

Vapour-phase equilibrium constants in the literature (determined, for example, from vapour density (i.e. *P-V-T*) measurements or from latent heats of vapourization) are available for a number of carboxylic acids as functions of temperature. However, the values reported in literature often show appreciable discrepancies, as shown for propanoic acid at two different temperatures in Table 2.1.

Table 2.1: Literature values for the vapour-phase chemical equilibrium constant of propanoic acid.

Literature Source	K_{AA} (kPa^{-1})	
	393.15 K	413.15 K
Tamir and Wisniak (1975), Kato et al. (1990)	0.02928	0.01144
Tsonopoulos and Prausnitz (1970)	0.03686	0.01541
Shih and Jones (1989)	0.04511	0.01629
Tamir and Wisniak (1978)	0.02989	0.01167

Accurate calculation of liquid-phase activity coefficients, therefore, becomes less reliable than for non-associating systems. This effect is amplified because both the calculation of the species concentrations (from Equations 2.12 and 2.13) and saturation pressures (Equation 2.17) depend on the equilibrium constants. An additional complication is that for the heterodimer constant, K_{AB}, a satisfactory basis must be found for its calculation from K_{AA} and K_{BB} for the pure carboxylic acids. Prausnitz et al. (1999), Nothnagel et al. (1973) and Christian (1957) assume a positive entropy change of $R\ln 2$ which leads to the expression:

$$K_{AB} = 2\sqrt{K_{AA}K_{BB}} \tag{2.20}$$

However, Campbell (1994) found that for alcohol mixtures that were modeled using a similar approach to that used in this work, the geometric mean

$$K_{AB} = \sqrt{K_{AA}K_{BB}} \tag{2.21}$$

produced good results. Interestingly, the radial distribution function analysis presented in Chapter 8 (Section 8.2.6.4) reinforces this empirical result. However, further study in this area (not possible during this project because of time constraints) is necessary before a definitive conclusion can be drawn.

Liquid-phase equilibrium constants (that would allow a more direct technique for calculating the liquid-phase activity coefficients, γ_i) are sparse in the literature. Barela et al. (1995) determined liquid-phase K values for substituted benzoic acids by a time-consuming isopiestic method. Sharma et al. (1983, 1999) measured freezing point depressions in dilute solvents and, from these, estimated molecular weights and K values, principally for coal-derived liquids and alcohols.

2.3 Thermodynamic Consistency

Although measured *P-T-x-y* data may be accurate, it is difficult to judge their thermodynamic consistency for associated systems, particularly since well-known tests depend on the Gibbs-Duhem equation (and, therefore, on calculated activity coefficients). These require an accurate description of the vapour phase, where non-

ideality may be principally or exclusively dependent on the association reactions and their respective equilibrium constants (which were shown to be rather uncertain in Section 2.2). Furthermore, reliable values for $P_{i_1}^{sat}$ (also subject to variations in the K_{ii} values) are required in Equation 2.19. The area test,

$$\int_{x_1=0}^{1} \ln\left(\frac{\gamma_1}{\gamma_2}\right) dx_1 = \int_{x_1=0}^{1} \frac{-H^E}{RT^2} dT + \int_{x_1=0}^{1} \frac{V^E}{RT} dP \tag{2.22}$$

though independent of the measured pressure and permitting a possible cancellation of errors, is nevertheless useful as a rapid indication of consistency and of the suitability of the postulated vapour phase reactions (Equations. 2.1 to 2.3). The excess volume for isothermal data and the excess enthalpy for isobaric data may reasonably be neglected for mixtures of chemically similar compounds (as is the case for systems consisting purely of carboxylic acids).

2.4 Bubble-point Calculation Algorithm

As was mentioned in the introductory paragraph of this Chapter, the T-K Wilson model (Tsuboka and Katayama, 1975) proved to be the most adept at modeling the liquid-phase activity coefficients of the carboxylic acid systems. This conclusion was reached based on the area test results (discussed further in Sections 8.1.2 and 8.1.3). The T-K Wilson expression for the liquid-phase activity coefficient in a binary system is (Walas, 1985):

$$\begin{aligned} \ln \gamma_i = \ln \frac{x_1 V_{i1} + x_2 V_{i2}}{x_1 \Lambda_{i1} + x_2 \Lambda_{i2}} + x_1 \left[\frac{V_{1i}}{x_1 + x_2 V_{12}} - \frac{\Lambda_{1i}}{x_1 + x_2 \Lambda_{12}} \right] \\ + x_2 \left[\frac{V_{2i}}{x_1 V_{21} + x_2} - \frac{\Lambda_{2i}}{x_1 \Lambda_{21} + x_2} \right] \end{aligned} \tag{2.23}$$

where $V_{ij} = V_j/V_i$, Λ_{ij} are defined as $V_{ij} = \exp[-(\lambda_{ji} - \lambda_{ii})/RT)]$ and $\Lambda_{ii} = 1$. The molar volumes (V_i) were calculated using the Rackett equation ($V_i = V_{ci} Z_{ci}^{(1-T_n)^{0.2857}}$). The T-K Wilson equation parameters $\Delta\lambda_{12}$ and $\Delta\lambda_{21}$ were fitted using the objective function recommended by Gmehling and Onken (1977) for the DECHEMA series:

$$O.F. = \sum \left(1 - \frac{\gamma_{calc}}{\gamma_{\exp}} \right)^2 \leq \varepsilon_\gamma \tag{2.24}$$

The bubble-point calculation procedure to find the system pressure P and the vapour composition, y_A, at any specified liquid composition, x_A, was developed as follows:

1. At the chosen T (within the experimental range) and x_A, the system pressure is estimated by $P = \sum_i x_i \gamma_i P_i^{sat}$ since the activity coefficients have already been determined using the T-K Wilson parameters (discussed above) and the vapour is assumed to be ideal for the initial step in the iteration scheme.

2. Calculate η_{A_1} and η_{B_1} from Equation 2.19.

3. Determine y_A and y_B from Equations 2.12 and 2.13 re-arranged:

$$y_A = \frac{2K_{AA}P\eta_{A_1}^2 + \eta_{A_1} + K_{AB}P\eta_{A_1}\eta_{B_1}}{K_{AA}P\eta_{A_1}^2 + 1 + K_{BB}P\eta_{B_1}^2 + K_{AB}P\eta_{A_1}\eta_{B_1}} \tag{2.25}$$

$$y_B = \frac{2K_{BB}P\eta_{B_1}^2 + \eta_{B_1} + K_{AB}P\eta_{A_1}\eta_{B_1}}{K_{BB}P\eta_{B_1}^2 + 1 + K_{AA}P\eta_{A_1}^2 + K_{AB}P\eta_{A_1}\eta_{B_1}} \tag{2.26}$$

4. Re-calculate P from the standard equation for the gamma-phi formulation of VLE:

$$P = \sum_i \frac{x_i \gamma_i P_i^{sat}}{\Phi_i} \tag{2.27}$$

with the vapour-phase correction for non-ideality, Φ_i, written in the following form:

$$\Phi_i = \frac{\eta_{i_1} P_i^{sat}}{y_i P_{i_1}^{sat}} \tag{2.28}$$

5. If $\left|P^{new} - P^{old}\right| > \varepsilon$ (ε was taken to be 10^{-3} kPa in this work) then return to Step 2 and continue iterating until the desired tolerance, ε, is attained.

The bubble-point calculation results are presented in Section 8.1.3.

3

Statistical Mechanics

Computer simulations are capable of providing very detailed information regarding the microstructure of physical systems. In addition, the position, orientation and (in the case of Molecular Dynamics) the motion of every molecule within the system is known and can be outputted for the researcher to analyze. However, this information is not extremely useful, since no laboratory experiment is able to produce equivalent data for comparison. Generally, an experiment measures an averaged property i.e. a macroscopic quantity (such as pressure, temperature or composition) averaged over all the particles (and often over time as well) comprising the system (Frenkel and Smit, 2002). To allow simulation data to be compared to this experimental data, similar averages of macroscopic properties are necessary. Statistical mechanics provides this information. Indeed, Hoover (1991) asserts that the "fundamental goal of Statistical Mechanics is to link the detailed determinism of many-body microscopic dynamics to the phenomenological averaged description of macroscopic behaviour".

3.1 Phase Space and System Ensembles

Before a discussion of statistical mechanics is begun, certain terms and phrases should be clearly defined to avoid any confusion that may arise at a later stage. The

paragraphs that follow provide a brief description of the fundamental expressions *phase space* and *ensemble*.

The phase space of a particular system is a multidimensional region consisting of phase points, each of which corresponds to a particular dynamical state of the system. Generally in statistical mechanics, the dynamical state of a system having F degrees of freedom may be definitively described by $2F$ dimensions; one dimension for each of the coordinates q_1 to q_F and each of the momenta p_1 to p_F that specify the state (Tolman, 1979). According to Hill (1956), it is noteworthy that when the system forces are given, specifying a phase point in the phase space of the system (at a particular time, t) will completely determine the past and future trajectories of that point as it moves according to the laws of mechanics through the phase space.

However, in the case of Monte Carlo simulation, time is not taken into account because the system is considered to be at equilibrium (unlike molecular dynamics in which time is, by definition, a crucial factor and particle velocities are calculated as the time is discretely stepped) and, therefore, only the coordinates of the constituent atoms (i.e. F dimensions) are required to define the phase space. In other words, for Monte Carlo simulation the phase space is made up simply of all the possible system positions or configurations. A simple illustration of phase space is provided by considering the casting of a die. On any particular throw, the die will come to rest in one of six possible orientations (differentiated by the number of dots visible on the uppermost face). Thus, for this system the phase space consists only of the six points corresponding to these potential orientations.

The thermodynamic state of a system is normally defined by specifying certain variables, for example: the number of molecules, N (in multicomponent systems, the number of molecules of each species, N_i, would be required), system temperature, T, and system pressure, P (Allen and Tildesley, 1987). Now, consider a large number of such systems that exist in isolation (i.e. are completely independent) and in different dynamical states, yet are all characterized by the same system variables (N, P and T) mentioned above (Hill, 1956). Thus, in essence, the various system states differ microscopically, but retain the same macroscopic properties. This set of related, yet independent microstates is referred to as an ensemble, and all systems forming this ensemble share the same phase space. This is true in this instance because the systems possess the same values for the variables N, P and T; however, more general ensembles exist where this is not the case (Hill, 1956).

3.2 The Hypothesis of Equal a priori Probabilities

The hypotheses of equal *a priori* probabilities and ergodicity (discussed in Section 3.1.3 below) must be introduced in order to allow application of statistical mechanics to the study of macroscopic thermodynamic systems. Essentially, these postulates are used to show that the experimentally observed property of interest equates to the ensemble average of that property (calculated using classical statistical mechanics). The argument leading to this conclusion is presented in Section 3.1.4.

The postulate of equal *a priori* probabilities does not have a general, rigorous proof and according to Tolman (1979) can only reasonably be justified through the correspondence between the postulate conclusions and the regular behaviour (revealed through empirical methods) of actual systems. Huang (1963) offers a clear, simple explanation of the postulate: when a macroscopic system is in thermodynamic equilibrium, its microstate is equally likely to be any microstate that satisfies the macroscopic conditions of the system. This postulate is generally justified, in classical statistical mechanics, by Liouville's theorem (a proof and discussion of the theorem may be found in the text of Huang (1963)).

3.3 The Ergodic Hypothesis

Tolman (1979) defines the ergodic hypothesis through the following statement: the phase point of any isolated system will pass successively through every point compatible with the system energy before returning to occupy its original position in the phase space. Basically, this means (for Monte Carlo simulation purposes) that every accessible point in the phase space may be reached from any other phase point in a finite number of Monte Carlo steps (Frenkel and Smit, 2002).

The postulate of equal *a priori* probabilities applies to systems that are members of a particular ensemble, viz. the microcanonical ensemble. This ensemble (amongst others) is discussed in Section 3.5 and is characterized by the following variables: N, V (volume) and E (energy). However, Hill (1956) asserts that the microcanonical ensemble leads to both the canonical and grand canonical ensembles and, therefore, the postulate of equal *a priori* probabilities may be assumed to be valid for these ensembles

as well. On the other hand, the postulate of ergodicity applies by definition to all ensembles. A fuller discussion of these hypotheses and the validity of statistical mechanics in obtaining accurate ensemble averages that effectively correspond with measurable thermodynamic properties may be found by consulting the excellent texts of Hill (1956), Huang (1963) and Tolman (1979).

3.4 Ensemble Averages

During a laboratory experiment, the measurement of a thermodynamic quantity is accomplished by observing the change of that property over a certain amount of time and then computing the average value. In other words, any observed property, A_{obs}, is actually a time average. Mathematically, this may be written as:

$$A_{obs} = \langle A \rangle_{time} = \frac{1}{\tau} \int_0^{\tau} A(\mathbf{p}, \mathbf{q}) dt \tag{3.1}$$

where $A(p,q)$ is the value of A for a particular microstate, τ is the total time of the experiment and $\mathbf{p}$ and $\mathbf{q}$ are the position and momentum vectors of the phase points making up the phase space of the system.

Equation 3.1 implicitly shows that the thermodynamic property of interest can be calculated using a time average over all the system states. This is prevented in practice by sheer numerical complexity and the magnitude of the undertaking (Hill, 1956). Therefore, an alternative method must be brought into play to determine this average value: the ensemble average.

An ensemble, as was described in Section 3.1 above, is a collection of systems distributed in a phase space. This distribution is characterized by a density probability function, $\rho_{pf}(p,q,t)$, defined in such a way that the number of representative points contained in the volume element $d\mathbf{p}.d\mathbf{q}$ located at (p,q) in the phase space at the instant t is equal to $\rho_{pf}(p,q,t).d\mathbf{p}.d\mathbf{q}$ (Huang, 1963). Since we are interested in ensembles that do not depend explicitly on time, the probability density function may be reduced to $\rho_{pf}(p,q)$ instead. Using this simplified expression for the probability function, the ensemble average of A is defined by (Huang, 1963):

$$\langle A \rangle_{ens} = \frac{\int A(\mathbf{p},\mathbf{q})\rho_{pf}(\mathbf{p},\mathbf{q})d\mathbf{p}d\mathbf{q}}{\int \rho_{pf}(\mathbf{p},\mathbf{q})d\mathbf{p}d\mathbf{q}} \tag{3.2}$$

It is clear from the form of Equation 3.2 that if a state is highly improbable (with a correspondingly small value of $\rho_{pf}(p,q)$), then it will contribute only slightly to the overall ensemble average. This is obviously a desirable outcome, since the value of A for an improbable microstate is likely to differ substantially from the actual value of the observed property that we are attempting to predict, thereby reducing the accuracy of the average being calculated.

Now, the postulates introduced in Sections 3.2 and 3.3 (and in particular the ergodic hypothesis) indicate that the average properties of the system may be obtained from a snapshot of the system, without the necessity of following the system's entire evolution through time. That is to say, the ensemble average is identical to the time average. Therefore, the observed property, A, may be determined by calculating the ensemble average over a sufficiently large number of individual microstates:

$$A_{obs} = \langle A \rangle_{time} = \langle A \rangle_{ens} \tag{3.3}$$

Naturally, the numerous microstates used to calculate the ensemble average will vary from one another to some extent. Ideally, these fluctuations should be small, with the value of A for any particular state deviating only slightly from the overall ensemble average. If this is not the case, then the ensemble is likely to be far from representative of the experimental system being studied (Hill, 1956) and many microstates will need to be sampled to obtain the correct thermodynamic averages. According to Huang (1963), a measure of this fluctuation may be obtained by analyzing the mean square deviation of the ensemble average:

$$\frac{\langle A^2 \rangle - \langle A \rangle^2}{\langle A \rangle^2} << 1 \tag{3.4}$$

If this equation is not satisfied (that is, if the mean square fluctuation is not small), then the validity of statistical mechanics (in that particular instance) should be questioned (Huang, 1963).

3.5 Common Ensembles

As has been discussed in the preceding sections, an ensemble is a collection of microstates that are used to estimate properties observed in real systems through calculation of an ensemble average. Obviously, an ensemble that suitably describes the system of interest is essential. To this end, several statistical mechanics ensembles are commonly used and will be described here. They are the microcanonical, canonical, isothermal-isobaric and grand canonical ensembles.

3.5.1 The Microcanonical Ensemble

The microcanonical ensemble, as was mentioned in Section 3.3, is characterized by the variables N, V and E being constant. The probability density function for this ensemble may be conveniently expressed proportionally (Allen and Tildesley, 1987):

$$\rho_{pf}(\mathbf{p},\mathbf{q}) \propto \delta(H(\mathbf{p},\mathbf{q}) - E) \tag{3.5}$$

where $H(\mathbf{p},\mathbf{q})$ is the Hamiltonian of the system and the δ function selectively chooses those states of a system consisting of N particles residing in a volume V with the desired energy, E. The Hamiltonian expresses the total energy of an isolated system as a function of the coordinates and momenta of the constituent particles and may be written as $H = K + U$, where K and U represent the kinetic and potential energy of the system, respectively (Frenkel and Smit, 2002).

The partition function is a relationship that reflects the statistical properties of a system in a state of equilibrium. Each ensemble discussed here is characterized by a particular partition function, which may be written for the microcanonical ensemble as follows (Allen and Tildesley, 1987):

$$Q = \sum_{\mathbf{p},\mathbf{q}} \delta(H(\mathbf{p},\mathbf{q}) - E) \tag{3.6}$$

This equation may be rewritten in the quasi-classical form, taking due care to conserve the particle indistinguishability of the summation by employing a factor of $1/N!$:

$$Q = \frac{1}{N!}\frac{1}{h^{3N}}\int \delta\big(H(\mathbf{p},\mathbf{q}) - E\big)d\mathbf{p}d\mathbf{q} \tag{3.7}$$

where h is Planck's constant. For each ensemble, a specific thermodynamic potential is used to relate the partition function to a fundamental thermodynamic quantity. In the case of the microcanonical ensemble, this quantity is the extensive property entropy:

$$S = k_B \ln Q \tag{3.8}$$

where S represents entropy and k_B is the Stefan Boltzmann constant.

3.5.2 The Canonical Ensemble

In the canonical ensemble, the variables N, V and T (temperature) are held constant and the probability density function is (Allen and Tildesley, 1987):

$$\rho_{pf}(\mathbf{p},\mathbf{q}) \propto \exp\left(\frac{-H(\mathbf{p},\mathbf{q})}{k_B T}\right) \tag{3.9}$$

The corresponding partition function is generally expressed as follows:

$$Q = \sum_{\mathbf{p},\mathbf{q}} \exp\left(\frac{-H(\mathbf{p},\mathbf{q})}{k_B T}\right) \tag{3.10}$$

and in quasi-classical form:

$$Q = \frac{1}{N!}\frac{1}{h^{3N}}\int \exp\left(\frac{-H(\mathbf{p},\mathbf{q})}{k_B T}\right)d\mathbf{p}d\mathbf{q} \tag{3.11}$$

For the canonical ensemble, the appropriate thermodynamic quantity to which the partition function is related is the Helmholtz free energy (Allen and Tildesley, 1987):

$$E_H = -k_B T \ln Q \tag{3.12}$$

with the Helmholtz free energy represented by E_H.

3.5.3 The Isothermal-Isobaric Ensemble

The ensemble presented here is characterized by the fact that in addition to the number of molecules (N), the pressure (P) and temperature (T) remain constant. The expression (Allen and Tildesley, 1987) for the isothermal-isobaric ensemble probability density function is similar to that of the canonical ensemble, differing merely by the addition of a PV term:

$$\rho_{pf}(\mathbf{p},\mathbf{q}) \propto \exp\left(\frac{-[H(\mathbf{p},\mathbf{q})+PV]}{k_B T}\right) \tag{3.13}$$

Allen and Tildesley (1987) note that for this ensemble the volume (V) has joined the microscopic particle vectors of position and momentum in comprising the state point. This adds slightly to the complexity of the associated partition function:

$$Q = \sum_{\mathbf{p},\mathbf{q}} \sum_{V} \exp\left(\frac{-[H(\mathbf{p},\mathbf{q})+PV]}{k_B T}\right) \tag{3.14}$$

The summation over the possible volumes that the ensemble microstates may occupy may also be rewritten as an integral in the quasi-classical form of the partition function. Allen and Tildesley (1987) observe that a basic unit of volume, V_o, must be selected in order to preserve the partition function in a dimensionless form. Mathematically, this is expressed as follows:

$$Q = \frac{1}{N!}\frac{1}{h^{3N}}\frac{1}{V_o}\int dV \int \exp\left(\frac{-[H(\mathbf{p},\mathbf{q})+PV]}{k_B T}\right) d\mathbf{p}d\mathbf{q} \tag{3.15}$$

The corresponding thermodynamic function is, in this case, the Gibbs free energy, E_G:

$$E_G = -k_B T \ln Q \tag{3.16}$$

Not surprisingly, the form of this relation is very similar to that found for the canonical ensemble, Equation 3.12.

3.5.4 The Grand Canonical Ensemble

The variables that are specified in the grand canonical ensemble are μ (chemical potential), V and T. The probability density function (Allen and Tildesley, 1987) is proportional to a term that is closely related to those of the previous two ensembles:

$$\rho_{pf}(\mathbf{p},\mathbf{q}) \propto \exp\left(\frac{-H(\mathbf{p},\mathbf{q})+\mu N}{k_B T}\right) \tag{3.17}$$

As can be seen from Equation 3.17, the number of molecules is also a variable in the grand canonical ensemble. This is expressed in the partition function as an additional summation over the possible system sizes (i.e. N):

$$Q = \sum_{\mathbf{p},\mathbf{q}} \sum_{N} \exp\left(\frac{-H(\mathbf{p},\mathbf{q})+\mu N}{k_B T}\right) \tag{3.18}$$

and rewriting in quasi-classical form yields (Allen and Tildesley, 1987):

$$Q = \sum_{N} \frac{1}{N!} \frac{1}{h^{3N}} \exp\left(\frac{\mu N}{k_B T}\right) \int \exp\left(\frac{-H(\mathbf{p},\mathbf{q})}{k_B T}\right) d\mathbf{p} d\mathbf{q} \tag{3.19}$$

Allen and Tildesley (1987) state that although it is occasionally useful to treat N as a continuous variable and integrate with respect to N, the sum over N is usually preferred. The appropriate thermodynamic expression to which the grand canonical partition function is related is $-PV/k_BT$, which gives us finally:

$$PV = k_B T \ln Q \tag{3.20}$$

3.5.5 Ensemble Transformations

The ensembles that have been presented in the preceding sections are essentially artificial constructs (Allen and Tildesley, 1987), introduced to allow calculation of macroscopic thermodynamic properties using averages determined through statistical mechanics. Ideally, therefore, they should produce equivalent ensemble averages for a

particular system. This has been found to be true (Fisher, 1964) in the thermodynamic limit (and for a system of infinite size) provided that phase transitions are avoided. A formal treatment of the procedure for transforming between ensembles is given by Hill (1956) and Lebowitz et al. (1967), and an outline of the method may be found in Allen and Tildesley (1987).

The advantage of this consistency between ensembles is that the thermodynamic properties of a system may be determined from whichever ensemble is most convenient. Allen and Tildesley (1987) provide an excellent description of how the basic thermodynamic quantities are calculated using the ensemble averages; with the relevant expressions listed below:

Internal energy:
$$E_{\text{int}} = \langle H(\mathbf{p},\mathbf{q}) \rangle = \langle K(\mathbf{p}) \rangle + \langle U(\mathbf{q}) \rangle \tag{3.21}$$

Kinetic temperature:
$$T_K = 2\langle K(\mathbf{p}) \rangle / 3Nk_B \tag{3.22}$$

Note that the average of the instantaneous kinetic energy function is equal to the system temperature, T (Allen and Tildesley, 1987).

Virial pressure:
$$P_v = \rho k_B T + \frac{W_v}{V} = \rho k_B T - \frac{1}{3V} \sum_i \sum_{j>i} r_{ij} \frac{dU(r_{ij})}{dr_{ij}} \tag{3.23}$$

where r_{ij} is the distance between molecular centers and W is defined as the internal virial, which is expressed mathematically in terms of the intermolecular pair virial function, $w(r_{ij})$:

$$w(r_{ij}) = r_{ij} \frac{dU(r_{ij})}{dr_{ij}} \tag{3.24}$$

Thermodynamic pressure:
$$P_T = \rho k_B T - \left\langle \frac{\partial U}{\partial V} \right\rangle \tag{3.25}$$

This expression for the thermodynamic pressure is developed from the expression (Hummer et al., 1998):

$$P_T = -\left(\frac{\partial E_H}{\partial V} \right)_{T,N} \tag{3.26}$$

According to Hummer et al. (1998), the potential energy (U) does not typically depend explicitly on the volume (V). In the absence of this dependence, $\partial U/\partial V$ becomes:

$$\frac{\partial U}{\partial V} = \frac{1}{3V}\sum_i \frac{\partial U}{\partial \mathbf{q}_i}.\mathbf{q}_i \tag{3.27}$$

indicating that (under these conditions) the expression for the thermodynamic pressure reduces to that of the virial (or mechanistic) pressure. It should be noted, however, that the internal energy does depend on the volume when an Ewald summation is used.

Excess chemical potential: $$\mu^{ex} = -k_B T \ln\left\langle \exp\left(\frac{-U_{test}}{k_B T}\right)\right\rangle \tag{3.28}$$

where U_{test} is the resultant change that would occur in the potential energy following the addition of a random particle to the system (Allen and Tildesley, 1987). This expression is valid for the canonical, isothermal-isobaric and grand canonical ensembles; however, a slightly different formula applies for the microcanonical ensemble and is given in Allen and Tildesley (1987).

3.5.6 Ensemble Applications

Although the classical ensembles discussed above should produce equivalent results, certain ensembles are better suited to particular applications. The isobaric-isothermal (constant NPT) ensemble is widely used in Monte Carlo simulations (Frenkel and Smit, 2002). This makes sense, considering that the majority of laboratory measurements occur under similar conditions. This ensemble may also be used to measure the equation of state of a model system (even if the virial pressure cannot be readily evaluated) and is often applied to systems that involve a first-order phase transition (Frenkel and Smit, 2002). The grand canonical ensemble (constant μVT) is used for simulations where information pertaining to the average number of particles (N) is of interest. A good example is adsorption studies, in which the grand canonical ensemble closely mimics the typical experimental conditions (Frenkel and Smit, 2002). Since the vast majority of experiments do not take place under constant NVE conditions, the microcanonical ensemble is seldom used. According to Frenkel and Smit (2002), the

ensemble is most frequently applied in situations where a poor random number generator is biasing the Boltzmann distribution or in the unlikely event that the exponentiation of the Boltzmann factor is computationally expensive because the microcanonical ensemble does not rely on the Boltzmann factor at all and, in addition, uses no random numbers in determining whether a trial move will be accepted or not.

3.6 The Radial Distribution Function

The radial distribution function gives the probability of finding a pair of atoms, i and j, a distance r_{ij} apart, relative to a completely random molecular distribution at the same density (Allen and Tildesley, 1987). It is a pair correlation function that describes, on average, how the atoms in a system are radially packed around one another (Côté et al., 2001). This function is typically designated using the symbol $g(r)$ and may be used to analyze the microstructure of the system being studied. The major benefit that the radial distribution provides is that information regarding $g(r)$ may be obtained from neutron and X-ray diffraction techniques (Frenkel and Smit, 2002). Thus, numerical results for $g(r)$ calculated from molecular simulation data may be compared to experimental values, permitting a quantitative assessment of the ability of the simulations to reproduce the microstructure observed experimentally. Mathematically, the radial distribution may be expressed as follows (Côté et al., 2001):

$$g(r) = \frac{N(r)}{4\pi\rho_{bulk} r^2 \Delta r} = \frac{\rho(r)}{\rho_{bulk}} \tag{3.29}$$

where $\rho(r)$ is the radial density, ρ_{bulk} is the bulk (or mean) system density and $N(r)$ is the mean number of atoms existing in a shell of width Δr, a distance r from the atom or site upon which the radial distribution is centered.

From the definition for the radial distribution function given above (Equation 3.29), it is clear that if the distribution of molecules happens to be completely random throughout the fluid, then $g(r)$ will assume a value of unity throughout. Any deviation of $g(r)$ from unity indicates that the molecules are interacting in some manner. This allows the researcher to discover whether, for example, hydrogen bonding is occurring in the system and also (to an extent) how widespread the phenomenon might be.

According to Allen and Tildesley (1987), when $g(r)$ is calculated from simulation data, a histogram is compiled of all pair separations that lie within a specific range (i.e. within a volume defined by Δr, a distance r from the atom or site of interest). However, the length (L) of the simulation box limits the range over which $g(r)$ may be calculated and, hence, to maintain consistency with the minimum image convention r must be less than or equal to $L/2$ (Allen and Tildesley, 1987). The radial distribution functions analyzed during this project are plotted and discussed in Section 8.2.6. An example of a radial distribution function for a Lennard-Jones fluid is shown in Figure 3.1 below. The value of the distribution function must reach a value of unity before the cut-off radius (set at a value of 14 Å throughout this work) is reached, and the potential truncation is applied. If this is not the case, then the accuracy and validity of the simulation results should be questioned.

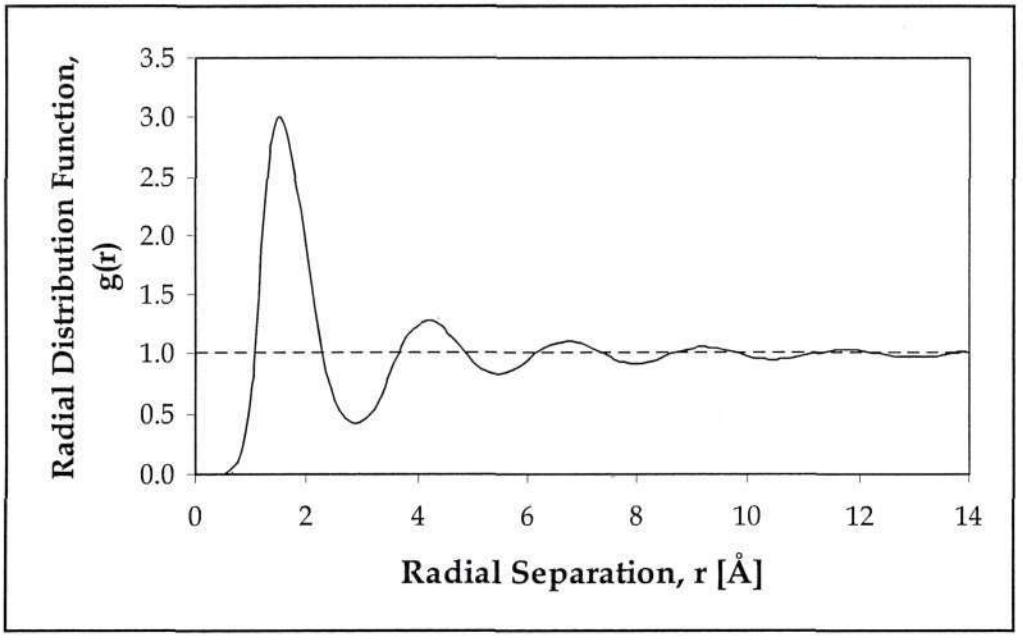

Figure 3.1: An example of a radial distribution function, $g(r)$, for a Lennard-Jones fluid.

4

Monte Carlo Simulation

Computer simulation of molecular systems has been feasible since the early 1950's. Metropolis et al. (1953) carried out the first simulation of a liquid and, simultaneously, introduced the technique that we know today as Monte Carlo simulation (Allen and Tildesley, 1987). Since the initial, pioneering work of Metropolis and co-workers the computing technology available to researchers has increased astronomically, allowing the study of ever more complex systems and molecules (e.g. proteins). This chapter provides an overview of the theoretical foundations of Monte Carlo simulation.

4.1 The Method of Monte Carlo Integration

The technique referred to as Monte Carlo integration was developed by von Neumann, Ulam and Metropolis in an effort to study neutron diffusion in fissionable material at the end of the Second World War (Allen and Tildesley, 1987). The name "Monte Carlo" was coined because of the widespread use of random numbers that the method demands. The invention of higher-order computers (namely, the MANIAC computer developed at Los Alamos) permitted the application of the Metropolis Monte Carlo method to the many-body problem that is unavoidably coupled to simulation of

molecular systems (Allen and Tildesley, 1987). A simple illustration of the benefit offered by Monte Carlo methods is provided by the following example:

Consider a researcher who wishes to calculate the amount of matter that exists in the observable universe. If conventional quadrature is used, then a predetermined (and, in this case, enormous!) array of equally spaced points is created and the quantity of matter that exists at each point is measured. A numerical technique such as Simpson's rule may then be used to obtain an estimate (from this data) as to the amount of matter in the observable universe. It is patently clear that an overwhelming proportion of these points will fall in areas where there is no matter whatsoever (i.e. in the vacuum of space). In other words, to obtain an accurate answer the mesh that is constructed must inevitably be very fine and, unfortunately, extraordinarily costly in computational terms. In the Metropolis Monte Carlo scheme, however, a random walk is constructed that travels through the relevant region of space, i.e. where the amount of matter present is nonnegligible (Frenkel and Smit, 2002). If a trial move attempts to sample an area of space that contains no matter, then this move will be rejected. Otherwise, the move is accepted and the process continues. After every trial move (accepted or not), the amount of matter present is recorded and the average of all these measurements eventually yields the final answer (Frenkel and Smit, 2002).

4.1.1 Random Sampling Method

The simplest form of Monte Carlo integration is known as random sampling. This technique may best be explained by studying the solution of a simple one-dimensional integral of a function $f(x)$ over the domain $x = [a, b]$:

$$I = \int_a^b f(x)dx \tag{4.1}$$

Obviously it is possible to use a conventional numerical technique to evaluate this integral. Another option, however, is available based on the unweighted average of $f(x)$ over the interval that extends from $x = a$ to $x = b$ (Frenkel and Smit, 2002). In this instance, Equation 4.1 may be expressed as:

$$I = (b-a)\langle f(x)\rangle \tag{4.2}$$

The required average of $f(x)$ is determined by evaluating the function at a large number of points randomly distributed between a and b (Frenkel and Smit, 2002). Provided that this sampling of the function is sufficiently extensive, the value calculated for the integral (I) will be extremely accurate. Furthermore, if the number of points is extended to infinity, then clearly I will be determined exactly. However, as was the case in the example described in Section 4.1, this method is extremely inefficient and a great deal of computing time is spent evaluating $f(x)$ at points where the function value is, in fact, negligibly small. Obviously, a procedure whereby many points are sampled in those regions in which $f(x)$ is significant and few elsewhere would be preferred (Frenkel and Smit, 2002). This method is called importance sampling and is the focus of the next section.

4.1.2 Importance Sampling Method

Importance sampling techniques are based on the selection of random numbers from a distribution that concentrates the function evaluation in those regions that make meaningful contributions to the integral (Allen and Tildesley, 1987), and the corresponding bias is corrected for later. This is accomplished by introducing a nonnegative probability density function, $\rho_{pf}(x)$ defined in a manner that ensures that the resultant sampling points are distributed nonuniformly over the integrand interval. To incorporate $\rho_{pf}(x)$ into Equation 4.1, we re-write the equation as follows:

$$I = \int_a^b \rho_{pf}(x) \frac{f(x)}{\rho_{pf}(x)} dx \tag{4.3}$$

Next, we make an assumption (Frenkel and Smit, 2002): we suppose that the function we have introduced is the derivative of a second (nonnegative, nondecreasing) function, $u(x)$, with $u(a) = 0$ and $u(b) = 1$. This permits I to be written as:

$$I = \int_0^1 \frac{f[x(u)]}{\rho_{pf}[x(u)]} du \tag{4.4}$$

with x expressed as a function of u. The function u is then used to generate random values (but uniformly distributed) between the bounds of the function (0 and 1). This

allows an estimate for I to be determined by summing over the total number of values generated from $u(x)$, ζ, and is calculated using Equation 4.5 below:

$$I = \frac{1}{\zeta} \sum_{i=1}^{\zeta} \frac{f[x(u_i)]}{\rho_{pf}[x(u_i)]} \tag{4.5}$$

The trick, of course, is to develop a probability density function that preferentially samples from those areas of the curve where $f(x)$ is nonnegligible, for the particular application of interest. Frenkel and Smit (2002) offer a more in-depth discussion of importance sampling methods and the variance that occurs as a result of the biased sampling scheme.

4.1.3 The Metropolis Method

The original Metropolis method (Metropolis et al., 1953) was developed for the canonical (i.e. constant NVT) ensemble. They demonstrated that it is possible to devise an efficient Monte Carlo scheme to sample a ratio of integrals of the form:

$$\langle A \rangle = \frac{\int \exp[-\beta U(\mathbf{q})] A(\mathbf{q}) d\mathbf{q}}{\int \exp[-\beta U(\mathbf{q})] d\mathbf{q}} \tag{4.6}$$

where $U(q)$ is the potential energy of configuration q and $\beta = 1/k_B T$. The ensemble average shown here differs from Equation 3.2 in that only the molecular positions are taken into account. This is always the case for molecular simulation Monte Carlo because the system is in a state of equilibrium and, therefore, the momenta ($\mathbf{p}$) do not form part of the phase space. In molecular dynamics computations, the momenta are clearly vital and must be included.

The principle difficulty encountered in applying the concept of importance sampling is discovering some means of generating a series of random states that ensures that each of those states has occurred (with the appropriate probability) by the end of the simulation (Allen and Tildesley, 1987). This is accomplished by constructing a Markov chain of states whose limiting distribution is the probability density function discussed in the previous section. According to Allen and Tildesley (1987), a Markov chain must satisfy two basic criteria:

1. Each trial has a finite set of possible outcomes, referred to as the state space.

2. The outcome of any particular trial depends exclusively on the outcome of the preceding trial.

Now consider any two arbitrary states that make up part of this chain. The probability of moving from the old state (designated as the configuration o) to the new state (n) may be written as $\pi(o \rightarrow n)$. When any Monte Carlo scheme is being developed, the principle of detailed balance should be adhered to. In essence, detailed balance (also known as microscopic reversibility) may be described using the expression below:

$$\rho_{pf}(o)\pi(o \rightarrow n) = \rho_{pf}(n)\pi(n \rightarrow o) \tag{4.7}$$

In words, this equation states that the average number of accepted moves from o to any other state n will be equivalent to the number of moves that occur in reverse i.e. from some state n to the state o (Frenkel and Smit, 2002). In reality, this condition of detailed balance is often unnecessarily strong (Allen and Tildesley, 1987) and the less stringent condition of balance is sufficient. This is desirable because it permits greater freedom in the selection and development of Monte Carlo moves. However, schemes based on the balance condition do not always result in correct Monte Carlo sampling and most algorithms that do not obey detailed balance are actually flawed (Frenkel and Smit, 2002).

There are many possible solutions for the probability matrix $\pi(o \rightarrow n)$ that will satisfy Equation 4.7 (Frenkel and Smit, 2002). The process of creating $\pi(o \rightarrow n)$ involves several related steps. First, a trial move is performed that shifts the system configuration from state o to state n. Once this move has been completed, the decision must be made whether to keep the move and accept the new state or not, with the probability of a trial move being accepted given by $\Xi(o \rightarrow n)$. The transition matrix that contains the probabilities of a certain move being performed between the two states is typically denoted by $\alpha(o \rightarrow n)$, and is also known as the underlying matrix of the Markov chain. This allows $\pi(o \rightarrow n)$ to be expressed in terms of the acceptance and transition matrices:

$$\pi(o \rightarrow n) = \alpha(o \rightarrow n) \times \Xi(o \rightarrow n) \tag{4.8}$$

Originally, Metropolis et al. (1953) chose α to be symmetric (Frenkel and Smit, 2002), which permits Equation 4.7 to be re-written in terms of acceptance matrices:

$$\rho_{pf}(o)\times\Xi(o\rightarrow n)=\rho_{pf}(n)\times\Xi(n\rightarrow o) \tag{4.9}$$

because $\alpha(o\rightarrow n)=\alpha(n\rightarrow o)$. Re-arrangement of Equation 4.9 yields:

$$\frac{\Xi(o\rightarrow n)}{\Xi(n\rightarrow o)}=\frac{\rho_{pf}(n)}{\rho_{pf}(o)} \tag{4.10}$$

The acceptance matrix $\Xi(o\rightarrow n)$ may be defined in various forms that satisfy Equation 4.10 (Frenkel and Smit, 2002). The Metropolis definition was:

$$\begin{aligned}\Xi(o\rightarrow n)&=\frac{\rho_{pf}(n)}{\rho_{pf}(o)} \quad \text{if} \quad \rho_{pf}(n)<\rho_{pf}(o)\\ &=1 \qquad\quad \text{if} \quad \rho_{pf}(n)\geq\rho_{pf}(o)\end{aligned} \tag{4.11}$$

which holds to the obvious condition that $\Xi(o\rightarrow n)$ must be less than or equal to unity. Interestingly, this original definition still seems to result in a more efficient sampling of the phase space than most strategies that have been proposed since (Frenkel and Smit, 2002).

The transition probability for the Metropolis scheme is, therefore:

$$\begin{aligned}\pi(o\rightarrow n)&=\alpha(o\rightarrow n) \qquad\quad \text{if} \quad \rho_{pf}(n)\geq\rho_{pf}(o)\\ &=\alpha(o\rightarrow n)\frac{\rho_{pf}(n)}{\rho_{pf}(o)} \quad \text{if} \quad \rho_{pf}(n)<\rho_{pf}(o)\end{aligned} \tag{4.12}$$

If a move is rejected, then the system must return to its original state. The probability of this occurring is, naturally, based on the probabilities of the new state being accepted:

$$\pi(o\rightarrow o)=1-\sum_{n\neq o}\pi(o\rightarrow n) \tag{4.13}$$

The acceptance criterion for the Metropolis method may therefore be summarized as follows:

$$\Xi(o\rightarrow n)=\min\left(1,\frac{\rho_{pf}(n)}{\rho_{pf}(o)}\right) \tag{4.14}$$

In practice, a move is accepted or rejected by generating a random number (from a uniform distribution between 0 and 1) and then comparing that random number to the acceptance probability of the attempted move i.e. to the ratio $\rho_{pf}(n)/\rho_{pf}(o)$. If the random number is less than $\Xi(o \rightarrow n)$, then the move will be accepted (Frenkel and Smit, 2002). A move that produces a ratio that is larger than unity will always be accepted. This makes sense because this generally occurs only when the new state, n, has a lower potential energy than the old state, o. This argument becomes clearer if one re-writes the probability ratio in terms of potential energy (Frenkel and Smit, 2002):

$$\frac{\rho_{pf}(n)}{\rho_{pf}(o)} = \exp\{-\beta[U(n)-U(o)]\} \tag{4.15}$$

If $U(n)$ is greater than $U(o)$, the term $-\beta[U(n)-U(o)]$ will be negative and, hence, the exponent of this value will be less than one. However, if $U(n)$ is less than $U(o)$, the term $-\beta[U(n)-U(o)]$ will be positive, the exponent will be greater than one and the move will be accepted. This ensures that moves that lead the system towards a state of thermal equilibrium are always accepted. Equation 4.15 is commonly termed a Boltzmann distribution.

It is evident that the method described above relies heavily on the random number generator and it is important that no bias be introduced into the procedure through the production of a non-uniform distribution. A problem arises because the numerical computations required to successfully run Monte Carlo simulations necessitate the use of computers in performing the calculations and, thus, the random numbers will also be generated by computer. However, computers are inherently deterministic and they are, therefore, generally incapable of producing a genuinely uniform random number distribution. Fortunately, current generators possess periods that are sufficiently long enough for the simulation to be completed (despite the huge quantity of random numbers required) before any repetition or bias occurs. Further details regarding random number generators may be obtained by consulting the following texts: James (1990), James (1994), Lüscher (1994) and Press et al. (1986).

In summary, then, the Metropolis method is a Markov process in which a random walk through phase space is constructed in a manner that results in the probability of visiting a particular system configuration, q, being proportional to the Boltzmann

factor i.e. $\exp[-\beta U(q)]$ (Frenkel and Smit, 2002). It should be noted, however, that only three of the classical ensembles introduced in the previous chapter use the Boltzmann distribution. The microcanonical ensemble acceptance criteria are based on a potential energy difference instead. For more details on the rarely used microcanonical ensemble, the reader is referred to the text of Frenkel and Smit (2002).

The Markov process random walk may be developed in various ways, but that put forward by Metropolis et al. (1953) was:

1. Randomly select a molecule and calculate the energy of that state $U(q_o)$.

2. Attempt to move the molecule a random distance in a random direction and calculate the energy associated with this new state $U(q_n)$.

3. Accept the move if the acceptance probability (Equation 4.14) is larger than or equal to a random number generated for comparison.

A detailed treatment describing the implementation of the Metropolis method in the common ensembles discussed in Section 3.5 is provided by Frenkel and Smit (2002). However, in this project the Gibbs ensemble was used for all simulations, and this successful and widely used ensemble is discussed in detail in Chapter 6.

4.1.4 Monte Carlo Trial Moves

Once the system under study has been set up in a suitable starting configuration and the potential model or models (discussed further in the next chapter) governing all molecular interactions have been specified, the underlying matrix of the Markov chain, α, must be decided (Frenkel and Smit, 2002). More specifically, this involves selection of the trial moves to be used. The particular trial moves that are implemented usually depend on the ensemble that has been chosen, but three broad categories may be identified:

- Volume perturbation (isothermal-isobaric ensemble (i.e. constant *NPT*) and other isobaric ensembles),

- Molecule translation, rotation and conformation (all ensembles),
- Molecule insertions and deletions (grand canonical ensemble i.e. constant μVT).

It should be noted that moves like the volume perturbations are not exclusively restricted to isobaric ensembles. For example, the Gibbs ensemble (as was mentioned above, this extremely useful ensemble is discussed further in Chapter 6) uses the volume move to achieve mechanical equilibrium via a volume exchange between the liquid and vapour boxes.

The actual design of a particular trial move is, for the most part, not restricted beyond the condition that detailed balance (or, at the very least, balance) must be maintained. This provides extensive scope for creativity in developing a move that enhances the sampling of the system phase space. Clearly, it is desirable for every state to be sampled according to its statistical probability during a simulation (and this would naturally occur in an infinitely long simulation). However, infinite simulations are obviously not a realistic possibility and, hence, trial moves that efficiently sample significant portions of the phase space are critical. This is more easily said than done, unfortunately, and it can often prove troublesome to encourage a system to adequately explore the full phase space and not merely sample extensively from an isolated region. A great deal of effort has been expended in the field of molecular simulation to design trial moves that satisfactorily sample complex systems. To this end, biased Monte Carlo techniques have been developed that greatly improve this process, and these methods are reviewed in the sections that follow.

4.2 Biased Monte Carlo Techniques

The methods discussed so far are fairly simple to implement for monatomic systems. However, chain molecules (or polymers) are frequently encountered in molecular simulation. These molecules introduce considerable complexity into the molecular translation, rotation, conformation and insertion moves, and the ratio of attempted to accepted moves becomes extremely tiny. For the Gibbs ensemble (which relies heavily on molecule exchanges between the two boxes), a sufficiently high acceptance rate for the insertion move is crucial. Usually, the chance of finding a suitable orientation for

the inserted molecule is low (especially for polar molecules). However, biased trial moves can be used to improve this acceptance (Frenkel and Smit, 2002).

The biased Monte Carlo techniques introduced here are largely based on Rosenbluth sampling (Rosenbluth and Rosenbluth, 1955). The chain conformations discussed below are constructed segment by segment (or atom by atom). Briefly, the Rosenbluth approach consists of two steps (Frenkel and Smit, 2002):

1. A chain conformation is created that is biased towards generation of "acceptable" conformations.

2. A weight factor (often termed the Rosenbluth factor) is used to correct the bias introduced in Step 1.

In other words, the Monte Carlo trial moves are no longer random, but are modified to enhance the probability that the attempted move will successfully fit into the existing configuration. Thus, the transition matrix is no longer symmetric and, to maintain detailed balance when using these biased moves, the acceptance rules must also be adjusted (Frenkel and Smit, 2002).

4.2.1 Orientational Bias

When simulations involve polar molecules and compounds that are capable of forming hydrogen bonds, the orientation of the molecules to one another becomes especially important. The challenge, therefore, develops into locating a position for the inserted molecule that does not overlap existing molecules, yet at the same time has an acceptable orientation relative to the adjacent molecules (Frenkel and Smit, 2002).

For the orientational bias trial move, we consider a random molecule that is to be moved and reoriented (Frenkel and Smit, 2002). As before, the old configuration is denoted o and the new configuration by n. For the translational part of the move, standard random displacement is used, but the generation of trial orientations is biased using the procedure laid out below (Frenkel and Smit, 2002).

First, the molecular center of mass (COM) is randomly displaced and all interactions independent of the molecules orientation ($U^{pos}(n)$) are calculated. A number of trial

orientations are generated $\{b_1, b_2, \ldots, b_k\}$ and the energy ($U^{or}(b_i)$) associated with each position is determined. The Rosenbluth factor for the new configuration is defined as:

$$W(n) = \sum_{j=1}^{k} \exp[-\beta U^{or}(b_j)] \tag{4.16}$$

The orientation of the selected molecule in the old configuration is denoted b_o. Again, various trial configurations are generated (in this case, k - 1 configurations) and the Rosenbluth factor for the old state may be calculated from Equation 4.17:

$$W(o) = \exp[-\beta U^{or}(b_o)] + \sum_{j=2}^{k} \exp[-\beta U^{or}(b_j)] \tag{4.17}$$

Finally, the move is accepted (as described for the Metropolis method in Section 4.1.3) based on the following probability:

$$\Xi(o \rightarrow n) = \min\left(1, \frac{W(n)}{W(o)} \exp\{-\beta[U^{pos}(n) - U^{pos}(o)]\}\right) \tag{4.18}$$

4.2.2 Standard Configurational Bias Monte Carlo

Configurational bias Monte Carlo (CBMC) owes its origins to the Rosenbluth scheme (Rosenbluth and Rosenbluth, 1955). However, the probability of generating a particular conformation using this scheme is not proportional to the Boltzmann factor (Frenkel and Smit, 2002). This shortcoming was recognized by Rosenbluth and Rosenbluth (1955) and the correction that they imposed was the Rosenbluth factor, W, mentioned in the previous section. The configurational bias scheme uses the Rosenbluth weighting factor to bias the acceptance of the trial conformations generated Rosenbluth method, and thereby guarantees that all chain conformations are generated with the correct Boltzmann weight (Frenkel and Smit, 2002).

The CBMC technique was developed by Siepmann and Frenkel (1992) and has become ubiquitous with simulation of molecular chains (i.e. polymers) because the algorithm significantly enhances the sampling for molecules with an articulated (i.e. branched) structure and improves the efficiency of chain insertions by several orders of

magnitude (Vlugt et al., 1998). A number of advanced algorithms have been developed from the original scheme and are discussed in later sections.

In the CBMC method, the potential energy is divided into two parts: the internal (or bonded) potential energy and the external potential energy. The internal energy that exists between two consecutive segments of a molecule is determined by summing the bond stretching, angle bending and the torsional potential energy contributions. All other energies (intermolecular and non-bonded intramolecular interactions) are tallied to yield the external potential energy, U^{ext}. This split is actually arbitrary and may be adjusted to suit a particular application or system.

The CBMC method may be used to grow a molecule from scratch, or alternatively, to partially regrow a molecule. An identical procedure is followed in both cases and will be outlined for a molecule that is to be re-grown. The technique begins with the random selection of a molecule chain, m, and a segment within that chain, j. All the segments with an index greater than j (or, just as likely, with an index lower than j) are discarded and the task then becomes to construct a conformation of l segments, starting from j and proceeding segment by segment until the molecule has been completely regrown or the trial move fails and the old configuration is retained (Siepmann and Frenkel, 1992).

Each segment is added in the same manner: a trial conformation, n, is created and thereafter the old configuration, o, is considered. To generate n, a number (k) of trial segments are created and (using similar notation to the previous section) are denoted $\{b_1, b_2, \ldots, b_k\}$. The probability of a certain trial segment, b, being generated is calculated from the bonded interactions of that segment (Frenkel and Smit, 2002). For every trial segment, the external Boltzmann factor ($\exp[-\beta U_i^{ext}(b_i)]$) is determined and a particular segment (now referred to as segment n) is chosen with a probability:

$$P_{gen} = \frac{\exp[-\beta U_i^{ext}(b_n)]}{\sum_{j=1}^{k} \exp[-\beta U_i^{ext}(b_j)]} \tag{4.19}$$

where the denominator is the Rosenbluth weight of segment i, $w_i(n)$ (Frenkel and Smit, 2002). When the entire chain has been re-grown, the Rosenbluth factor for the chain is calculated as a product of the Rosenbluth weights of the inserted segments:

$$W(n) = \prod_{i=1}^{l} w_i(n) \tag{4.20}$$

The Rosenbluth factor of the old configuration, $W(o)$, must now be determined using a similar procedure (Frenkel and Smit, 2002). First, a chain is randomly selected. The Rosenbluth weight of the first bead is:

$$w_1(o) = k \exp[-\beta U_1^{ext}(o)] \tag{4.21}$$

where k retains its definition as the number of trial segments created earlier in the scheme. The Rosenbluth weights of the remaining segments are given by:

$$w_i(o) = \sum_{j=1}^{k} \exp[-\beta U_i^{ext}(b_j)] \tag{4.22}$$

The k orientations that are being summed in Equation 4.22 are constructed as follows (Frenkel and Smit, 2002): $k - 1$ orientations are generated in exactly the same manner as before, and the actual bond between segment i and segment $i - 1$ completes the set. The Rosenbluth factor for the entire chain may now be determined:

$$W(o) = \prod_{i=1}^{l} w_i(o) \tag{4.23}$$

The final step is to determine whether to accept the move, based on the acceptance probability (Frenkel and Smit, 2002):

$$\Xi(o \to n) = \min\left(1, \frac{W(n)}{W(o)}\right) \tag{4.24}$$

Naturally, for a molecule that is to be grown from scratch i.e. from the first bead (e.g. if a molecule is to be inserted into a system), the initial steps of selecting segment j and discarding units from the molecule are not necessary. In this instance, the scheme begins with the insertion of segment one, and then proceeds in the manner described above.

4.2.3 Coupled-decoupled Configurational Bias Monte Carlo

The standard CBMC method works well for linear molecules. However, branched compounds present more of a challenge. At first glance, it might appear that the chain could still be grown successfully using the standard technique discussed above. Vlugt et al. (1999) proved that this is not, in fact, the case and were able to show that the angle distributions generated are not correct. In other words, sequential construction of the trial orientations for a branched molecule introduces a bias. Naturally, this bias could be removed in the acceptance criterion, but may produce an inefficient algorithm (Frenkel and Smit, 2002). This bias could simply be ignored, and Vlugt et al. (1999) state that the error incurred is only significant at high temperatures. Clearly, however, it would be preferable to remove as many sources of error as possible. The couple-decoupled CBMC technique developed by Martin and Siepmann (1999) does exactly that by introducing nested CBMC loops.

Martin and Siepmann (1999) demonstrated that the problem revealed by Vlugt et al. (1999) may be overcome through the rather simple expedient of including the internal energies when performing the biased growth of a molecule (normally based on the external potential energy, U^{ext}, only):

$$P_{gen} = \prod_{i=1}^{l} \left[\frac{\exp[-\beta U_i^{ext}(b_n)] \exp[-\beta U_i^{\text{int}}(b_n)]}{\sum_{j=1}^{k} \exp[-\beta U_i^{ext}(b_j)] \exp[-\beta U_i^{\text{int}}(b_j)]} \right] \tag{4.25}$$

The primary advantage of this scheme is that it permits decoupling of the selection of the different energies; the ensuing benefit is that a large number of trial sites may be chosen for the less computationally expensive bond angle selection, without increasing the cost of performing the other biased selections (Martin and Siepmann, 1999). The probability of generating a given configuration is:

$$P_{gen} = \prod_{i=1}^{l} \left[\frac{\exp[-\beta U_i^{LJ}(b_L)]}{W_L(i)} \right] \left[\frac{\exp[-\beta U_i^{tors}(b_T)]}{W_T(i)} \right] \left[\frac{\exp[-\beta U_i^{ang}(b_A)]}{W_A(i)} \right] \tag{4.26}$$

and the corresponding Rosenbluth weights for each energy selection are:

$$W_L(i) = \sum_{b_L=1}^{k_L} \exp[-\beta U_i^{LJ}(b_L)] \tag{4.27}$$

$$W_T(i) = \sum_{b_T=1}^{k_T} \exp[-\beta U_i^{tors}(b_T)] \tag{4.28}$$

$$W_A(i) = \sum_{b_A=1}^{k_A} \exp[-\beta U_i^{ang}(b_A)] \tag{4.29}$$

where U_i^{LJ} is the potential energy from the Lennard-Jones interactions ($U_i^{LJ} = U_i^{ext} + U_i^{bond}$) and k_L, k_T and k_A are the number of trial sites for the Lennard-Jones, torsional and angle bending interactions, respectively.

The main drawback of the method, according to Martin and Siepmann (1999), is that once the torsional angles have been selected in a biased fashion, there is only one possible trial site available to the Lennard-Jones selection. Their solution to this problem is to couple the biased selections, obtaining:

$$P_{gen} = \prod_{i=1}^{l} \left[\frac{\exp[-\beta U_i^{LJ}(b_L)] W_T(b_L)}{W_L(i)} \right] \left[\frac{\exp[-\beta U_i^{tors}(b_T)] W_A(b_T)}{W_T(b_L)} \right] \left[\frac{\exp[-\beta U_i^{ang}(b_A)]}{W_A(b_T)} \right] \tag{4.30}$$

$$W_L(i) = \sum_{b_L=1}^{k_L} \exp[-\beta U_i^{LJ}(b_L)] W_T(b_L) \tag{4.31}$$

$$W_T(b_L) = \sum_{b_T=1}^{k_T} \exp[-\beta U_i^{tors}(b_T)] W_A(b_T) \tag{4.32}$$

$$W_A(b_T) = \sum_{b_A=1}^{k_A} \exp[-\beta U_i^{ang}(b_A)] \tag{4.33}$$

The move is accepted with probability:

$$\Xi(o \to n) = \min\left(1, \frac{\prod_{i=1}^{l} W_L^n(i)}{\prod_{i=1}^{l} W_L^o(i)}\right) \tag{4.34}$$

The consequence of this biased selection coupling is that for each trial step of the Lennard-Jones biased selection, a separate torsional bias selection is performed (Martin and Siepmann, 1999), alleviating the problem mentioned above. However, the coupled selection also carries a disadvantage: the number of trial vectors for the angle bending bias selection becomes very large ($k_L.k_T.k_A$ vectors). Hence, in general, a combination of coupled and decoupled selections is used; providing considerable flexibility in the design of a CBMC scheme for a specific system or molecule (Martin and Siepmann, 1999). The onus, therefore, is on the researcher to devise an appropriate scheme with the number of trial vectors chosen to ensure adequate sampling without unnecessary computational expense.

4.2.4 Dual Cutoff Configurational Bias Monte Carlo

The dual cutoff CBMC (DC-CBMC) scheme was developed to improve the efficiency of trial orientation generation and Rosenbluth weight calculation in a CBMC simulation (Vlugt et al. 1998). Vlugt et al. (1998) demonstrated that use of the DC-CBMC algorithm resulted in a three to four fold improvement in computation times (for a system of 144 octane molecules). Furthermore, the speedup arising from the use of a center of mass (COM) based cutoff was found to improve as system size increased. For octane molecules, Vlugt et al. (1998) found that the optimal inner cutoff radius, r_{cut}*, is approximately 4 Å. This is more or less the diameter of an alkane pseudo-atom, and indicates that the excluded volume is the dominant term in the CBMC growth biasing selection.

Consider two molecules M_1 and M_2 that are representative of the simulation system. Both molecules possess a maximum distance from its COM to any interaction site denoted d_{M1} and d_{M2}, respectively. If the cutoff radius (discussed in more detail in the next chapter) is r_{cut}, the molecules cannot have any direct interaction if their separation extends beyond a distance of d_{M1} + d_{M2} + r_{cut} (Vlugt et al. 1998). Therefore, all interaction site distances do not need to be determined (Martin and Siepmann, 1998), unless a molecule conformation change is successfully accepted. Vlugt and co-workers showed that this COM cutoff decreases CPU time by as much as five times for a system of 1152 octane molecules. Now, this COM cutoff can also be implemented when calculating the Boltzmann weights of the trial orientations for a particular segment, i

(Vlugt et al., 1998). A list is compiled of all molecules whose separation from i is less than the combined distance calculated by summing the bond length between beads i and i - 1, the separation of j from i (d_{Mj}) and r_{cut}. To compute the Rosenbluth weight of the conformation, the external energy of each bead comprising these molecules must be determined for each trial site. It is evident that the list of molecules can quickly become extensive and the number of computations required extremely large. The introduction of the inner cutoff radius in the DC-CBMC technique is intended to address this issue.

The algorithm begins by splitting the Lennard-Jones potential energy into two parts:

$$U^{LJ} = \overline{U}^{LJ} + \delta U^{LJ} \tag{4.35}$$

where $\bar{U}^{LJ}$ is a potential that is less expensive to calculate than U^{LJ} and δU^{LJ} is (fairly obviously) the difference between these two potentials (Vlugt et al. 1998). Many choices exist for $\bar{U}^{LJ}$, but DC-CBMC uses an inner cutoff radius, r_{cut}^* to significantly reduce the list of molecules that interact with segment i. This leads to the expression:

$$U^{LJ}(r) = \overline{U}^{LJ}(r < r_{cut}^*) + \delta U^{LJ}(r_{cut}^* < r < r_{cut}) \tag{4.36}$$

To ensure that the correct distribution is recovered, the acceptance probability requires some modification (Vlugt et al. 1998):

$$\Xi(o \to n) = \min\left(1, \frac{\overline{W}(n)}{\overline{W}(o)} \exp\left\{-\beta[\delta U^{LJ}(n) - \delta U^{LJ}(o)]\right\}\right) \tag{4.37}$$

where the Rosenbluth weights in Equation 4.37 are calculated for the potential $\bar{U}^{LJ}$. Thus, for a DC-CBMC re-growth the full potential is not determined for each trial conformation, but only for the final configuration. The proof for Equation 4.37 is given in Vlugt et al. (1998), as well as proof that the DC-CBMC distribution is equivalent to the distribution generated by the standard CBMC scheme.

4.2.5 Aggregation-volume-bias Monte Carlo

Strongly associating compounds are very important in many industrial sectors, and this was one of the primary reasons for the investigation of these species in this project.

The simulation of systems involving these compounds can be extremely complex due to the polar interactions that are occurring (for example, hydrogen bonding). These interactions can often lead to phenomena such as the formation of molecular clusters. For systems in which clusters are likely to form, adequate phase space sampling becomes difficult (without resorting to extremely long simulation times) and the simulation frequently becomes trapped in a particular bonded configuration. The aggregation-volume-bias Monte Carlo (AVBMC) algorithm was designed to improve the efficiency of this sampling.

The scheme was developed by Chen and Siepmann (2000), according to whom the AVBMC method is easy to implement, generally applicable, robust and offers several advantages over competing algorithms. AVBMC enhances the acceptance rate of trial moves that result in molecular clusters breaking up, and simultaneously increases the transition probability of moves that lead to the formation of bonded structures (Chen and Siepmann, 2000). It is obvious that this will improve the sampling of the phase space of systems in which such configurations are likely to form, provided that the algorithm works as intended. Chen and Siepmann (2000) were able to demonstrate that this is indeed the case and that their algorithm greatly increases the likelihood that a molecule will be added to (or removed from) a particular cluster.

An AVBMC trial move proceeds in the following manner (Chen and Siepmann, 2000): two molecules (say i and j) are randomly selected, with $i \neq j$. The second molecule, j, serves as a target molecule for the swap move involving i. This swap move occurs based on a probability, P_{swap}: the move is restricted to the bonded region of molecule j (termed the n_{in} state) with probability P_{swap}, and to the non-bonded region of j (the n_{out} state) with probability $1 - P_{swap}$.

The crucial part of the AVBMC algorithm is the acceptance probabilities for the four possible scenarios that exist, viz. $o_{out} \rightarrow n_{out}$, $o_{out} \rightarrow n_{in}$, $o_{in} \rightarrow n_{out}$ and $o_{in} \rightarrow n_{in}$. Two of the cases ($o_{out} \rightarrow n_{out}$ and $o_{in} \rightarrow n_{in}$) do not involve i entering or leaving the bonded region of j and, therefore, the standard acceptance rules (like those discussed in the preceding sections) apply. For the second case, $o_{out} \rightarrow n_{in}$, the acceptance probability is (irrespective of whether i was bonded to a molecule other than j before the move was attempted or not):

$$\Xi(o_{out} \rightarrow n_{in}) = \min\left(1, \frac{(1-P_{swap})V_{in}\exp[-\beta(E_n - E_o)]}{P_{swap}V_{out}}\right) \tag{4.38}$$

where V_{in} is the volume of the bonded region about molecule j and V_{out} is the remainder of the volume existing in the system (i.e. $V_{out} = V - V_{in}$). If the trial move removes i from a bonded configuration with j (and irrespective of whether i ends up bonded to another molecule again or not), the move is accepted with probability:

$$\Xi(o_{in} \rightarrow n_{out}) = \min\left(1, \frac{\mathrm{P}_{swap} V_{out} \exp[-\beta(E_n - E_o)]}{(1 - \mathrm{P}_{swap}) V_{in}}\right) \tag{4.39}$$

The underlying Markov matrix is asymmetric for the AVBMC method, with the bias that is introduced proportional to P_{swap}/V_{in} and $(1 - \mathrm{P}_{swap})/V_{out}$ for trial moves placed into the *in* and *out* regions of j, respectively (Chen and Siepmann, 2000). The bias of the underlying Markov matrix, in conjunction with the acceptance criteria (Equations 4.38 and 4.39) may be substituted into Equations 4.7 and 4.8 to show that the scheme satisfies the detailed balance condition discussed earlier in the chapter (Section 4.1.3).

5

Molecular Interactions

Any molecular simulation consists of a system of molecules that interact with one another. These interactions are controlled, in general, by the model that calculates the inter- and intra-molecular potential energies. The aim, of course, is to replicate as closely as possible the conditions that would exist in a real system (or, at the very least, to generate the same values or trends for the thermodynamic properties of interest). Obviously, the number of particles that exist in a laboratory experiment cannot be matched by a computer simulation and imaginative and ingenious methods must be applied to approximate the real system. This chapter introduces the techniques used to achieve this goal, within the limits imposed by the available technology and theory.

5.1 Periodic Boundary Conditions

As was noted above, molecular simulations attempt to reproduce the properties of a real thermodynamic system based on averages collected from a small number of molecules that are representative of the system. In such small systems, the majority of the molecules are situated on the surface of the container (or box) and they experience appreciably different conditions to those that lie in the bulk. Allen and Tildesley (1987) describe how for a system consisting of 1000 molecules, arrayed in a 10×10×10 cube,

almost 49 % of the molecules (488) are located on the cube faces. The number of molecules used in this project did not exceed 800 and, hence, any surface effects would have significantly affected the results. The implementation of periodic boundary conditions (Born and von Karman, 1912) provides an elegant means of avoiding this complication.

In essence, the technique creates a box in which all the molecules that form the system are located. This box does not possess any walls or boundaries as such, and merely serves as a convenient means of storing the coordinates of the molecules (Allen and Tildesley, 1987). An infinite system possessing the characteristics of the bulk fluid is formed by replicating this cubic box through space. As the molecules move in the original box during the course of a simulation, the molecules in the replicated boxes move in an identical fashion. Thus, if a molecule passes through any side of the central box, a molecule simultaneously enters the box from the corresponding surrounding box. The method is illustrated in Figure 5.1 below:

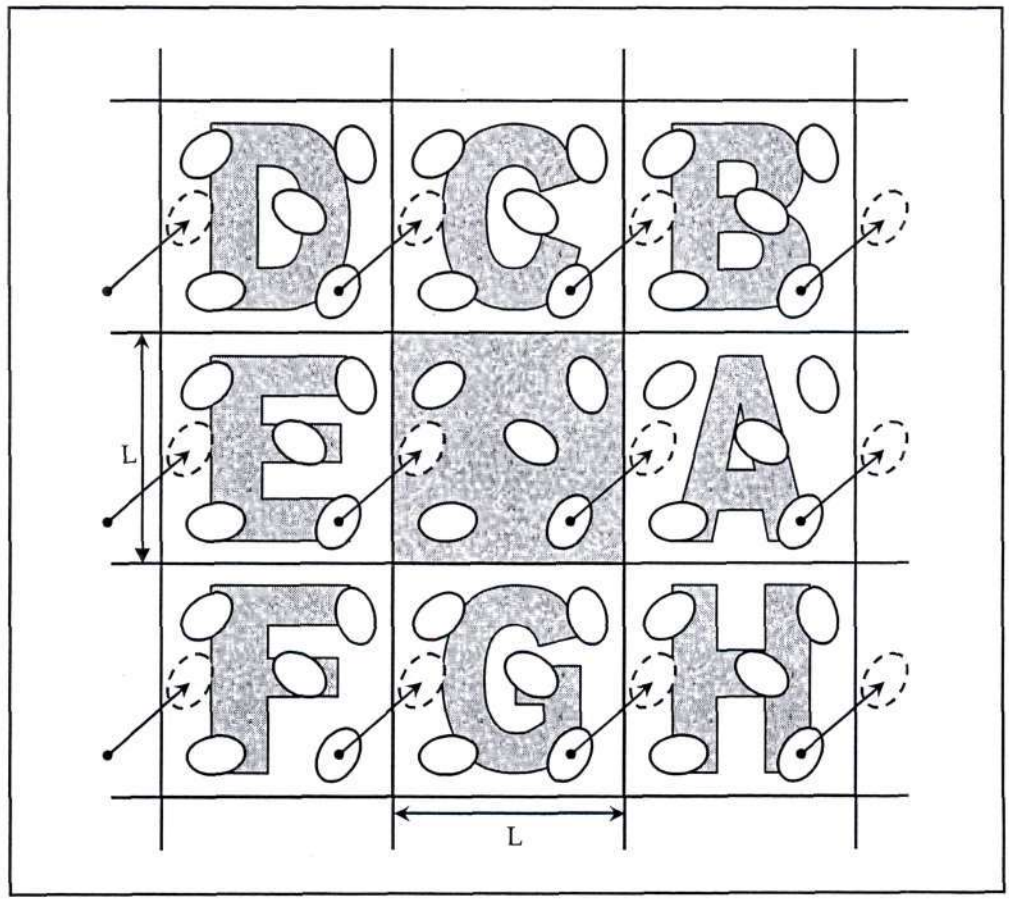

Figure 5.1: The method of periodic boundary conditions (Allen and Tildesley, 1987).

Figure 5.1 shows a two-dimensional system rather than the three-dimensional situation that would exist in an actual simulation. This simplification is adopted for the sake of clarity. It is evident that boxes A to H are identical to the central box in every respect. This has two obvious advantages: first, the fact that the particles in each image move in an identical fashion means that the number density in the central box is always conserved (Allen and Tildesley, 1987). Second, it eases the computational demand since only the coordinates of the molecules in the primary image need to be recorded.

Each molecule in the infinite system represented above will interact with all other molecules in the system, including its own periodic image. If the assumption is made that all intermolecular interactions in the system are pairwise-additive, then the total potential energy of the bulk fluid would be (Frenkel and Smit, 2002):

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

where N is the number of molecules in each periodic box, $\mathbf{r}_{ij}$ is the distance separating molecules i and j, $\mathbf{n}$ is an arbitrary vector of three integer numbers and L is the box length (see Figure 5.1). It should be noted the when $\mathbf{n} = 0$, the $i = j$ term should be excluded from the sum (Frenkel and Smit, 2002).

However, the infinite sum (Equation 5.1) is of little practical use in calculating the bulk behaviour. Fortunately, the intermolecular interactions are usually meaningful only over rather small distances (i.e. they are mostly short-range in nature) and the potential may be truncated beyond a predetermined cutoff radius. Potential truncation is discussed further, later in the chapter (Section 5.3).

Although the technique of periodic boundary conditions is extremely useful for simulating bulk systems, there are certain drawbacks that should be considered before the method is applied. The longest fluctuation wavelength that can be accommodated using periodic boundary conditions is $\lambda = L$ (Frenkel and Smit, 2002). Thus, any simulation in which fluctuations with a wavelength longer than the box length (L) are anticipated will encounter problems. Allen and Tildesley (1987) and Frenkel and Smit (2002) offer several examples of systems in which such fluctuations are important. Another shortcoming of the method is that the molecules may be able to "sense" the symmetry of the periodic lattice if the boxes are created too small (Allen and Tildesley, 1987). Also, long-range potentials (such as those necessary for simulation of polar compounds) result in significant interaction between a molecule and its images in the

surrounding boxes. The technique employed to deal with these potentials is addressed in Section 5.4. Finally, periodic boundary conditions may sometimes induce anisotropies in the fluid structure of certain systems (Allen and Tildesley, 1987; Frenkel and Smit, 2002).

5.2 Potential Models

Potential models (also known as potential energy surfaces or force fields) calculate the potential energy of the system and are used to determine all of the molecular interactions that occur in a simulation. The models are commonly divided into two parts: the intramolecular and intermolecular contributions. The intramolecular interactions, as implied by the name, occur between the atoms that comprise individual molecules. For organic molecules like hydrocarbons, the hydrogen atoms are often "collapsed" onto the carbon atoms to form so-called pseudo-atoms, drastically reducing the number of interactions that must be computed. Intermolecular interactions, on the other hand, occur between different molecules and are long-range forces, sometimes extending beyond the periodic box boundaries.

5.2.1 Intramolecular Interactions

The intramolecular potential is generally divided into the following components:

1. The potential energy due to bonded interactions (e.g. bond stretching), U^{bond}.
2. The potential energy resulting from angle bending, U^{ang}.
3. The torsional potential energy, U^{tors}.
4. The potential energy arising from non-bonded interactions, $U^{nonbond}$.

Each contribution to the intramolecular potential is calculated separately, and then summed to determine the total internal potential energy, U^{int}. The separate energy contributions are discussed in greater detail in the sections that follow.

5.2.1.1 *Bonded Interactions*

The potential energy contributed by the bonded interactions is typically represented by a harmonic potential that permits bond stretching:

$$U^{bond} = \frac{1}{2}\sum_{bonds} k_b (b - b_0)^2 \tag{5.2}$$

where k_b is the harmonic constant for bond stretching, b is the bond length and b_0 is the equilibrium bond length. The energy calculated for each bond in the system is summed to form the overall bond stretching potential energy, U^{bond}. The equilibrium bond length may be determined experimentally or, alternatively, computationally using an accepted scientific package such as the Gaussian 98 suite of subroutines (Frisch et al., 1998).

The use of this harmonic potential (Equation 5.2) provides flexibility to the model, although certain force fields use fixed bond lengths instead. Other functions for representing atomic bonds also exist; for example, Morse bonds (Morse, 1929).

5.2.1.2 *Angle Bending*

Any molecule that possesses three or more atoms or sites (and is not forced to remain rigid i.e. a constant bond angle) will have an energy contribution from the bending of the angles formed between the various atoms. As for the bond stretching discussed above, a harmonic potential is used to describe this interaction:

$$U^{ang} = \frac{1}{2}\sum_{angles} k_\theta (\theta - \theta_0)^2 \tag{5.3}$$

where k_θ is the harmonic constant for angle bending, θ is the bond angle, θ_0 is the equilibrium bond angle and U^{ang} is the total angle bending potential energy. The equilibrium bond angle is determined in the same manner as the equilibrium bond length.

5.2.1.3 *Torsional Potential*

The torsional potential is used to calculate the energy fluctuations that occur as the dihedral angle of a molecule is varied. Only molecular chains consisting of at least four atoms or sites are able to form dihedral angles (ϕ). Figure 5.2 illustrates the dihedral angle (for a butane molecule) and an example of a torsional energy curve is shown in Figure 5.3. The dihedral angle is defined as the angle created between the planes formed by atoms A, B, C and B, C, D (see Figure 5.2).

The torsional potential energy is an even function that is typically described using a series of cosine functions. The various torsional potential functions used in the TraPPE force field are given in Equations B.1 to B.3 of Appendix B, with the relevant constants presented in Table B.3. As may be observed in Figure 5.3, the potential energy minimum corresponds to the *trans* form of the butane molecule. This makes sense because the steric hindrance between the two CH3 pseudo-atoms is minimized for this particular molecular configuration. The *trans* conformer occurs at $\phi = 0°$ in Figure 5.3, but this is an arbitrary designation and could as easily have been situated at $\phi = 180°$ (i.e. with the *cis* conformer located at $\phi = 0°$).

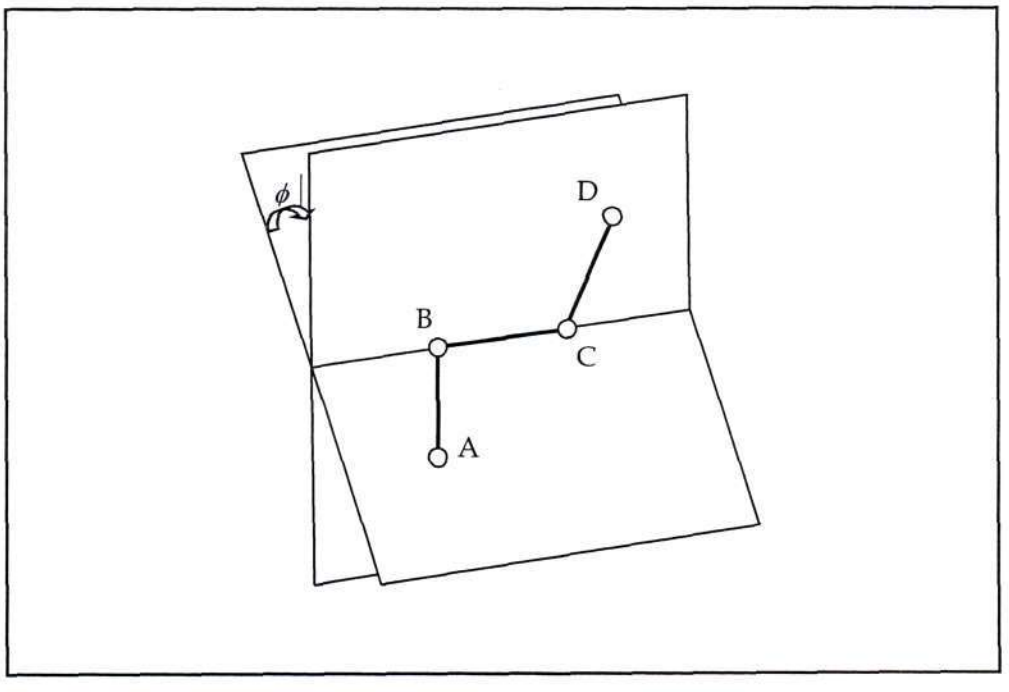

Figure 5.2: Illustration of the dihedral angle of the butane molecule (Allen and Tildesley, 1987).

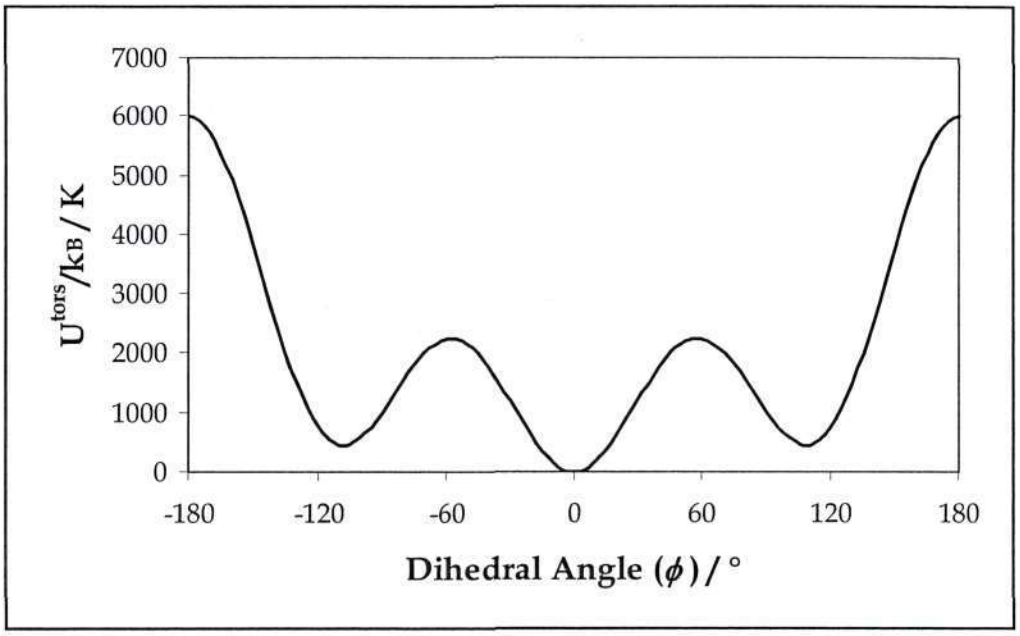

Figure 5.3: Example of a torsional potential energy curve (not for any particular molecule).

5.2.1.4 *Non-bonded Interactions*

The potential energy that arises as a result of non-bonded intramolecular interactions is calculated in exactly the same manner as for intermolecular interactions, although a scaling factor may be introduced. Thus, for polar compounds, the non-bonded potential has both a van der Waals and a Coulombic component. This is discussed further in Section 5.2.2 below. The intramolecular non-bonded interactions occur in molecules that contain pseudo-atoms separated by more than three bonds. Hence, this potential is often referred to as the "1-4 non-bonded interaction".

5.2.2 Intermolecular Interactions

As was mentioned in the previous section, the intermolecular potential consists of a van der Waals and a Coulombic function for polar molecules. Clearly, the polar and non-polar interactions are described by the Coulombic and van der Waals functions, respectively.

5.2.2.1 *The van der Waals Function*

The functional forms that have been used by researchers to express the van der Waals interactions are many and varied. Some examples (Allen and Tildesley, 1987) include the hard-sphere potential (Equation 5.4), the soft-sphere potential (Equation 5.5) and the square-well potential (Equation 5.6):

$$U^{HS}(r_{ij}) = \begin{cases} \infty & (r_{ij} < \sigma) \\ 0 & (\sigma \le r_{ij}) \end{cases} \tag{5.4}$$

$$U^{SW}(r_{ij}) = \begin{cases} \infty & (r_{ij} < \sigma_1) \\ -\varepsilon & (\sigma_1 \le r_{ij} < \sigma_2) \\ 0 & (\sigma_2 \le r_{ij}) \end{cases} \tag{5.5}$$

$$U^{SS}(r_{ij}) = \varepsilon\left(\frac{\sigma}{r_{ij}}\right)^{\upsilon} \tag{5.6}$$

where ε and σ are the energy and size parameters for pair potentials and υ is a parameter commonly chosen to be an integer.

However, the most common van der Waals function is the Lennard-Jones potential. Both the Lennard-Jones 12-6 and 9-6 functional forms are frequently employed (where the numbers refer to the power indices used in the Lennard-Jones expression), with the 12-6 form being used throughout this project:

$$U^{LJ}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{5.7}$$

The Lennard-Jones potential actually consists of two components: a repulsive part and an attractive part, with the equilibrium separation taken to occur at the energy minimum. This may be expressed as follows (Allen and Tildesley, 1987):

$$U^{RLJ}(r_{ij}) = \begin{cases} U^{LJ}(r_{ij}) + \varepsilon_{ij} & (r < r_{\min}) \\ 0 & (r_{\min} \le r) \end{cases} \tag{5.8}$$

$$U^{ALJ}(r_{ij}) = \begin{cases} -\varepsilon_{ij} & (r_{ij} \le r_{\min}) \\ U^{LJ}(r_{ij}) & (r_{\min} \le r_{ij}) \end{cases} \tag{5.9}$$

The repulsive part dominates at very short intermolecular distances (i.e. when the electron clouds of the two molecules are in close proximity) and the attractive part begins to dominate as the intermolecular separation increases. This is illustrated in Figure 5.4 where the attractive and repulsive Lennard-Jones functions are plotted.

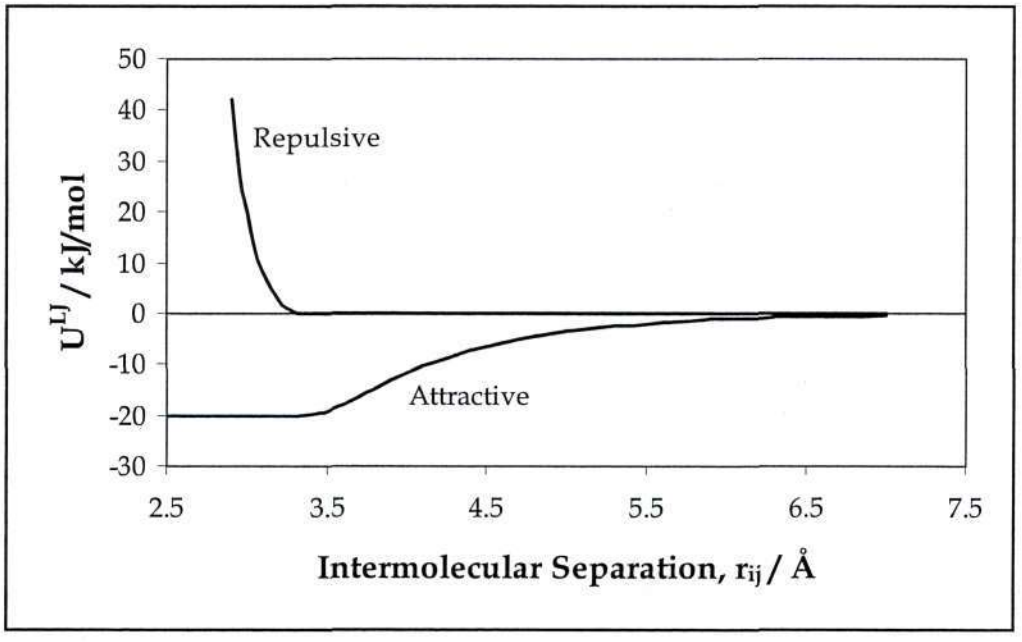

Figure 5.4: The repulsive and attractive Lennard-Jones 12-6 functional forms.

The combined Lennard-Jones 12-6 potential is plotted in Figure 5.5 below. The Lennard-Jones energy (ε_{ij}) and size (σ_{ij}) parameters used in Figures 5.4 and 5.5 are 20 kJ/mol and 3 Å, respectively. The minimum potential energy occurred at an intermolecular separation of approximately 3.4 Å. At the beginning of the Lennard-Jones functional form discussion, the 9-6 form of the expression was mentioned. In this case, the repulsive portion of the Lennard-Jones function is not as "hard" i.e. the slope is not as steep as for the 12-6 form. This provides a better representation of the van der Waals repulsion for certain molecules and molecules interacting with surfaces.

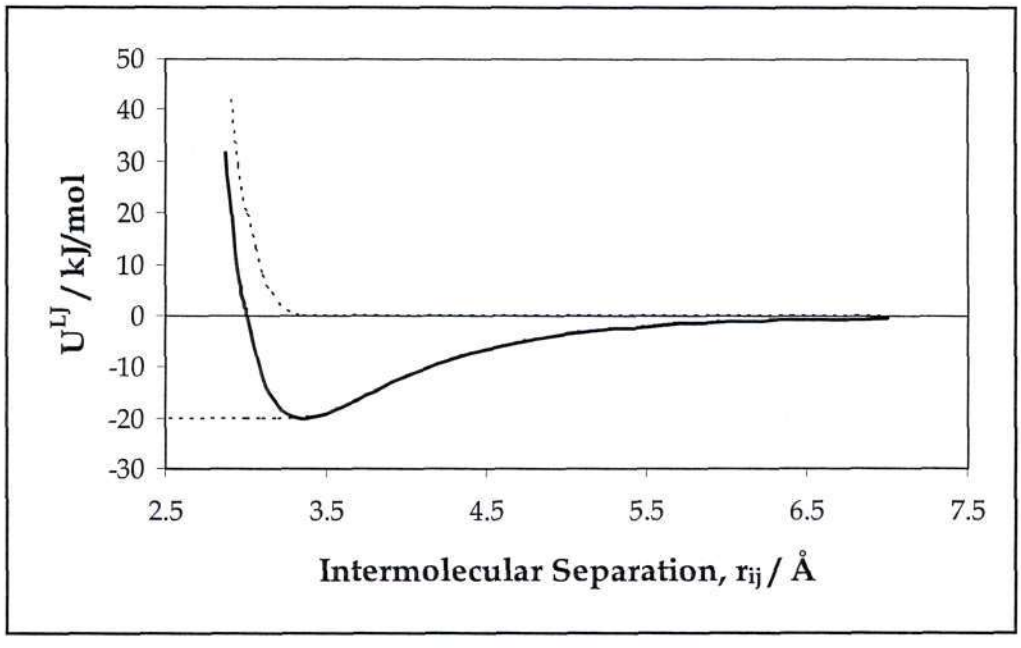

Figure 5.5: The combined Lennard-Jones 12-6 functional form (the repulsive and attractive components are represented by dashed lines).

As computing power improves, more complicated (and, hence, more accurate and computationally demanding) models are constantly becoming available. The Lennard-Jones 12-6 potential energy description is a fairly simple expression that works surprisingly well for many different compounds, both polar and non-polar. However, polar compounds require an additional, Coulombic term to correctly describe the intermolecular interactions.

5.2.2.2 *Coulombic Interactions*

For polar compounds, the van der Waals potential is not sufficient to represent the long-range, charged interactions (Allen and Tildesley, 1987). The simplest solution is to introduce a Coulombic interaction potential:

$$U^{charge}(r_{ij}) = \frac{q_i q_j}{4\pi r_{ij}} \tag{5.10}$$

where q_i and q_j are the atomic charges of sites i and j. The Ewald summation technique (discussed in Section 5.4 below) is used to calculate the potential energy contribution of the charged interactions.

5.2.3 Combining Rules

Consider a pure component system consisting of Z atom (or pseudo-atom) types. For this system, ½Z(Z + 1) sets of Lennard-Jones energy and size parameters must be specified to permit computation of the van der Waals interactions. For small molecules (e.g. H_2S or SO_2) this amounts to 3 parameter sets, but for a molecule like 2-methyl propanoic acid, 21 parameter sets are necessary. It is obvious that the number of parameter sets required rises rapidly as the size and complexity of the compound being simulated increases. Clearly, this problem becomes even more pronounced in multicomponent systems.

To minimize the number of parameters that must be explicitly declared, combining rules (sometimes called mixing rules) are used to calculate the unlike binary interaction parameters from the like parameters determined during the parameterization of a particular force field. In fact, the mixing rules should be defined and used in the parameterization, but are often used when combining force fields that have been developed separately for different molecules. It should be noted that the calculation of unlike interaction parameters in a simulation containing compounds described by different force fields is only possible if the force fields share the same van der Waals function (e.g. Lennard-Jones 12-6). The combining rules used during this project are introduced briefly in the sections that follow.

5.2.3.1 *Lorentz-Berthelot*

The Lorentz-Berthelot (Lorentz, 1881; Berthelot, 1898) combining rule is very popular and widely used. Arithmetic and geometric means are used for the size (σ) and energy (ε) parameters, respectively:

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \tag{5.11}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \tag{5.12}$$

The Lorentz-Berthelot rule has no physical basis and, although this combining rule is frequently used by researchers, it has been shown that it is not always optimal (or even suitable) for systems containing polar compounds (discussed further in Section 8.2.5).

5.2.3.2 *Geometric*

Like the Lorentz-Berthelot mixing rule, the geometric combining rule is empirical. As the name suggests, both the energy and size unlike-pair parameters are estimated using a geometric mean:

$$\sigma_{ij} = \sqrt{\sigma_{ii}\sigma_{jj}} \tag{5.13}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \tag{5.14}$$

5.2.3.3 *Kong*

The Kong combining rule (Kong, 1973) is an improvement over the Lorentz-Berthelot and geometric rules in that it was developed according to the atomic distortion model i.e. it has a physical basis (Delhommelle and Millié, 2001). Originally, the Kong mixing rule was proposed in an effort to improve the agreement between simulation and

experimental data for noble-gas mixtures (Potoff et al., 1999). The Lennard-Jones parameters are calculated using the following simultaneous expressions:

$$\varepsilon_{ij}\sigma_{ij}^{12} = \frac{\varepsilon_{ii}\sigma_{ii}^{12}}{2^{13}}\left[1+\left(\frac{\varepsilon_{jj}\sigma_{jj}^{12}}{\varepsilon_{ii}\sigma_{ii}^{12}}\right)^{\frac{1}{13}}\right]^{13} \tag{5.15}$$

$$\varepsilon_{ij}\sigma_{ij}^{6} = \sqrt{(\varepsilon_{ii}\sigma_{ii}^{6})(\varepsilon_{jj}\sigma_{jj}^{6})} \tag{5.16}$$

5.2.3.4 *Waldman-Hagler*

This combining rule was proposed for rare gases by Waldman and Hagler (1993). They found that an arithmetic combining rule for σ_{ij}^6 (rather than for just σ_{ij}) produced accurate results for the unlike-pair diameter ((Delhommelle and Millié, 2001). The Waldman-Hagler mixing rule is:

$$\sigma_{ij} = \sqrt[6]{\frac{\sigma_{ii}^{6}+\sigma_{jj}^{6}}{2}} \tag{5.17}$$

$$\varepsilon_{ij} = \frac{\sqrt{\varepsilon_{ii}\sigma_{ii}^{6}\varepsilon_{jj}\sigma_{jj}^{6}}}{\sigma_{ij}^{6}} \tag{5.18}$$

5.2.3.5 *Halgren*

The Halgren combining rule (Halgren, 1992) was, like the Waldman-Hagler rule, proposed (and tested) for rare gases. This combining rule differs substantially from those presented in the preceding sections, with the most unusual feature being the expression to determine ε_{ij}: the equation is the harmonic mean of its own harmonic and geometric means (Halgren, 1992). The unlike interaction size parameter, on the other hand, is calculated using a cubic mean relationship. This mixing rule was proposed because the arithmetic and geometric rules consistently produce a minimum energy

distance (σ_{ij}) that is too small when the like-pair distances differ substantially (Halgren, 1992). Halgren (1992) also found that the geometric mean rule for well-depths leads to large departures from experimental values and consistently over-estimates the unlike-pair well-depth, particularly when the like-pair parameters are markedly dissimilar.

$$\sigma_{ij} = \frac{\sigma_{ii}^3 + \sigma_{jj}^3}{\sigma_{ii}^2 + \sigma_{jj}^2} \tag{5.19}$$

$$\varepsilon_{ij} = \frac{4\varepsilon_{ii}\varepsilon_{jj}}{\left(\sqrt{\varepsilon_{ii}} + \sqrt{\varepsilon_{jj}}\right)^2} \tag{5.20}$$

5.3 Potential Truncation

Truncation of the intermolecular potential was briefly mentioned in Section 5.1, and is possible because the potential is short-range in nature, i.e. the greatest contribution to the potential comes from neighbours that are close to the molecule of interest (Allen and Tildesley, 1987). This supposition is necessary because of the infinite lattice created when periodic boundary conditions are applied. Each molecule should, in theory, interact with every molecule that exists in the infinite system; clearly an impossible task. To reduce the problem to manageable proportions, the first step is to apply the assumption known as the "minimum image convention".

5.3.1 Minimum Image Convention

In the minimum image convention, each molecule is considered to be at the center of a region that is exactly the same size and shape as the basic simulation box shown in Figure 5.1 (Allen and Tildesley, 1987). The molecule then interacts only with the surrounding molecules whose centers fall within this locale. Figure 5.6 (on the facing page) provides a graphical illustration of the minimum image convention.

Broadly speaking, therefore, the minimum image convention is a basic form of potential truncation where the truncation boundary is the image (or periodic box)

surrounding the molecule of interest. Although, this assumption makes calculation of the potential energy feasible, the pairwise-additive interactions still involve $\frac{1}{2}N(N - 1)$ terms (Allen and Tildesley, 1987). Depending on the total number of molecules in each box (N), this may still be a very large (and time-consuming) computation. A further approximation (the simple potential truncation) improves the situation considerably.

5.3.2 Simple Potential Truncation

The simple potential truncation is implemented by setting the pair potential to zero outside of a user-defined spherical region, i.e. $U^{LJ}(r_{ij}) = 0$ for $r_{ij} \geq r_{cut}$, where r_{cut} is the cutoff radius (Allen and Tildesley, 1987). This spherical cutoff is represented by the dashed circle centered on molecule 5 in Figure 5.6. The only molecules that contribute to the truncated potential are those whose centers fall within the sphere defined by $r_{5j} < r_{cut}$. Thus, the interactions between molecule 5 and molecules 2^C and 4^E will be included, but interactions with molecules 1 and 3^D will be excluded. According to Allen and Tildesley (1987), for a cubic simulation box with box length L, the number of surrounding molecules that must be considered is reduced by a factor of $4\pi r_{cut}^3/3L^3$.

Clearly, the computational saving increases as r_{cut} decreases and, therefore, the cutoff radius should be as small as possible. However, it is vital that the truncation only introduce a minor perturbation to the system energy. Typically, the radius is chosen to be approximately two and a half times the Lennard-Jones size parameter (σ). For the Lennard-Jones potential (Equation 5.7), the potential at $r_{cut} = 2.5\sigma$ is only 1.6 % of the well depth (Allen and Tildesley, 1987). It should also be noted that to comply with the minimum image convention, we must have $r_{cut} < \frac{1}{2}L$. The radial distribution function (refer to Section 3.6) provides a useful means of testing whether the cutoff radius is large enough: the function ought to reach a value of unity (indicating that the radial density is equal to the bulk density) before the cutoff radius is reached. If this is not the case, then r_{cut} should be increased.

The potential truncation assumption is clearly a powerful tool for improving simulation speeds. However, the truncation process introduces a certain amount of error, and the thermodynamic properties will differ from those calculated using the full potential (Allen and Tildesley, 1987). Fortunately, long-range corrections may be applied to rectify the situation.

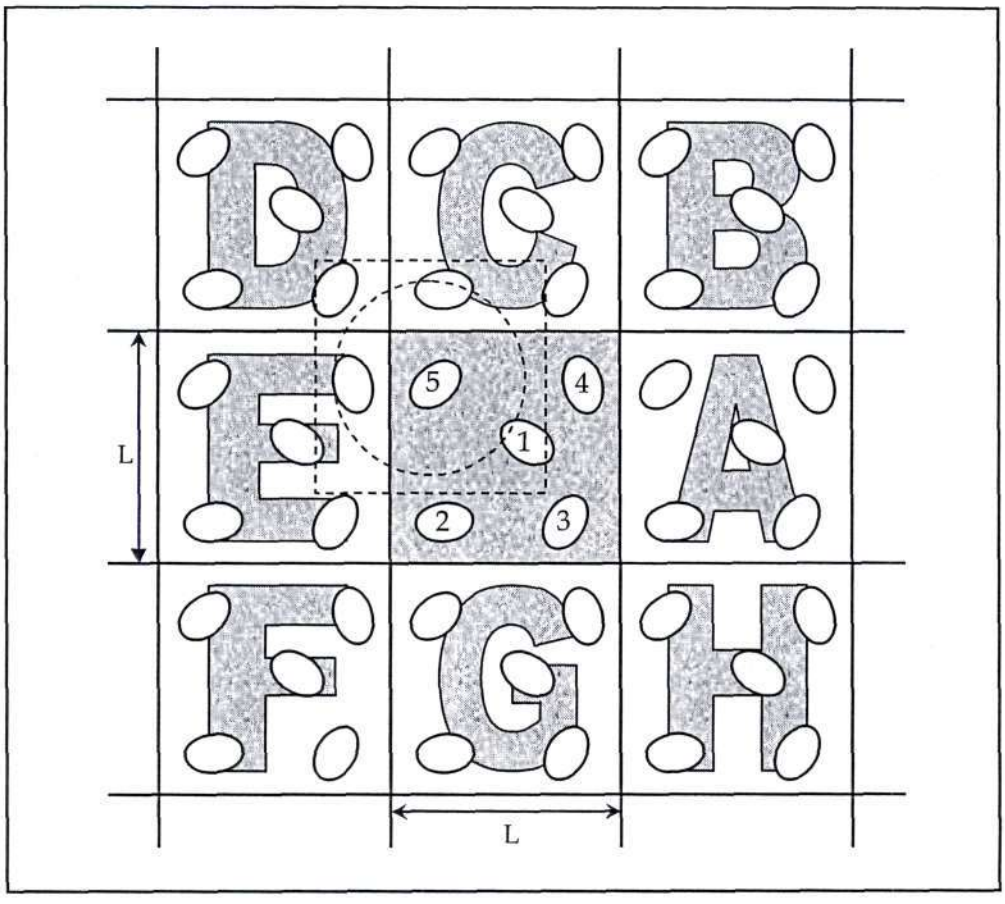

Figure 5.6: Illustration of the minimum image convention (Allen and Tildesley, 1987). The dashed box is the region representing the minimum image and the dashed circle shows the spherical potential cutoff.

5.3.3 Analytical Tail Correction

Frenkel and Smit (2002) state that the average potential energy of any atom i may be calculated from:

$$U_i = \frac{1}{2}\int_0^\infty 4\pi r^2 \rho_n(r) U(r) dr \tag{5.21}$$

where $\rho_n(r)$ is the average number density at a given distance, r, from i. Truncating the potential at r_{cut} means that the tail contribution of the integral given in Equation 5.21 is ignored. This may be expressed mathematically (Frenkel and Smit, 2002):

$$U_i^{tail} = \frac{1}{2}\int_{r_{cut}}^{\infty} 4\pi r^2 \rho_n(r) U(r) dr \tag{5.22}$$

The analytical solution of Equation 5.22 is simplified somewhat by assuming that $\rho_n(r)$ is equal to the average number density, ρ_n, for $r \geq r_{cut}$ (Frenkel and Smit, 2002). For the Lennard-Jones 12-6 potential model, Equation 5.22 becomes:

$$U_i^{tail} = 8\pi\rho_n\varepsilon\int_{r_{cut}}^{\infty} r^2\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]dr \tag{5.23}$$

Equation 5.23 reduces, upon reduction of the integral, to the following:

$$U_i^{tail} = \frac{8}{3}\pi\rho_n\varepsilon\sigma^3\left[\frac{1}{3}\left(\frac{\sigma}{r_{cut}}\right)^{9} - \left(\frac{\sigma}{r_{cut}}\right)^{3}\right] \tag{5.24}$$

It is obvious that the derivation above is for a pure Lennard-Jones 12-6 fluid, i.e. all the atoms are identical and σ and ε are constant for every atom in the system. The situation is more complicated when different pseudo-atoms are involved, but the procedure for determining the tail corrections remains unchanged.

5.3.4 Shifted Potential Truncation

The technique of not only truncating, but simultaneously shifting the potential is normally only used in Molecular Dynamics simulations. The shifted potential is necessary to prevent the introduction of an intermolecular potential discontinuity and, consequently, the intermolecular forces are always finite (Frenkel and Smit, 2002). This is important because certain Molecular Dynamics algorithms (based on Taylor expansions) cannot handle impulsive forces (Frenkel and Smit, 2002). The truncated and shifted potential is expressed as follows:

$$U_i^{LJ-shift}(r_{ij}) = \begin{cases} U^{LJ}(r_{ij}) - U^{LJ}(r_{cut}) & (r_{ij} \leq r_{cut}) \\ 0 & (r_{ij} > r_{cut}) \end{cases} \tag{5.25}$$

5.3.5 Hard-inner Cutoff Radius

On occasion, a trial move will attempt to insert a pseudo-atom into a region that either overlaps or is in close proximity (separated by a distance of less than 2 Å) to an existing molecule. The resulting system configuration will obviously have an extremely high energy (especially in the case of molecular overlap) and, consequently, will be rejected. To save computational time, a hard-inner cutoff radius may be used instead. Any move that places a pseudo-atom nearer than this cutoff radius (typically set to a value of approximately 1 Å, but may range anywhere between 0 and 2 Å depending on the nature of the system in question) from an existing atom is automatically rejected. In general, simulations involving polar molecules require a smaller hard-inner cutoff radius because the attractive Coulombic interactions subdue the repulsive van der Waals forces that exist when molecules are brought into close proximity. This allows the molecules to lie somewhat closer to one another than would be the case in a non-polar system.

5.4 The Ewald Summation

The Coulombic potential (Equation 5.10) that is used to calculate the energy resulting from charged interactions in a system of polar compounds was introduced in Section 5.2.2.2. However, this is by definition a long-range force ($U^{charge}(r_{ij}) \sim r_{ij}^{-1}$) and its influence extends well beyond the limits of the periodic box. Thus, a technique is required that is capable of handling such a potential. Several possible solutions exist, but the most common (and the method used in this project) is the Ewald sum.

Consider a system consisting of N particles (both positively and negatively charged, but of equal number i.e. the system is electrically neutral overall) in a cubic box of length L and volume V with periodic boundary conditions enforced. The Ewald sum provides a means of efficiently computing the interactions between the charged particles and all of their periodic images. The potential energy may be expressed as follows (Allen and Tildesley, 1987):

$$U^{charge} = \frac{1}{8\pi} \sum_{\mathbf{n}=0}^{\infty} \sum_{i,j=1}^{N} \frac{q_i q_j}{\left|\mathbf{r}_{ij} + \mathbf{n}L\right|} \tag{5.26}$$

and, as for Equation 5.1, when $\mathbf{n} = 0$, the $i = j$ term is excluded from the sum. Unfortunately, Equation 5.26 is conditionally convergent (i.e. the order in which the terms are added together effects the final result) for long-range potentials (Allen and Tildesley, 1987). A convenient choice is to base the summation order on the proximity of the periodic boxes to the central simulation box. Thus, the infinite lattice is constructed of (approximately) spherical layers radiating outwards from the central box located at $\mathbf{n} = (0, 0, 0)$. This approach, however, necessitates that the relative permittivity, ε_s, of the medium surrounding the sphere be specified (Allen and Tildesley, 1987). Typical medium choices are a good conductor e.g. a metal ($\varepsilon_s = \infty$) or a vacuum ($\varepsilon_s = 0$), but the final results (although closely related) differ slightly. Further details regarding the effect of the medium (and the Ewald summation in general) may be found by consulting the texts of de Leeuw et al. (1980a), de Leeuw et al. (1980b) and Heyes (1981).

The charges that are used in the Ewald sum method are point charges, and the electrostatic energy is, as was mentioned above, a poorly converging sum. However, the electrostatic potential of a set of screened charges is far easier to calculate by direct summation because it is a rapidly decaying function of r_{ij} (Frenkel and Smit, 2002). Hence, to improve the convergence of Equation 5.26, we screen each point charge with a diffuse charge distribution of opposite sign that exactly cancels the point charge, q_i (Frenkel and Smit, 2002). Generally, a simple Gaussian distribution is selected for this screening charge. To compensate for the addition of the screening charge cloud to the system, a second, oppositely-charged, smoothly-varying Gaussian charge distribution is included (see Figure 5.7 below).

The compensatory charge distribution is advantageous because it is a smoothly varying and periodic function, and may, therefore, be conveniently represented by a rapidly converging Fourier series (Frenkel and Smit, 2002). At the site of each point charge, therefore, the overall electrostatic potential consists of three parts: the contribution of the point charge, the screening charge cloud and the compensatory charge cloud. Naturally, any Coulombic self-interactions should be excluded, but to retain the beneficial properties of the canceling distribution, the contribution from this charge distribution is included and corrected for later (Frenkel and Smit, 2002).

The screening and compensatory charge distributions are both taken to be Gaussian, and are conveniently expressed in the following form (Allen and Tildesley, 1987):

$$\chi_i(\mathbf{r}) = q_i\left(\frac{\alpha}{\pi}\right)^{3/2} \exp\left(-\alpha r^2\right) \tag{5.27}$$

where α is an arbitrary parameter that determines the distribution width and $\mathbf{r}$ is the position relative to the center of the distribution.

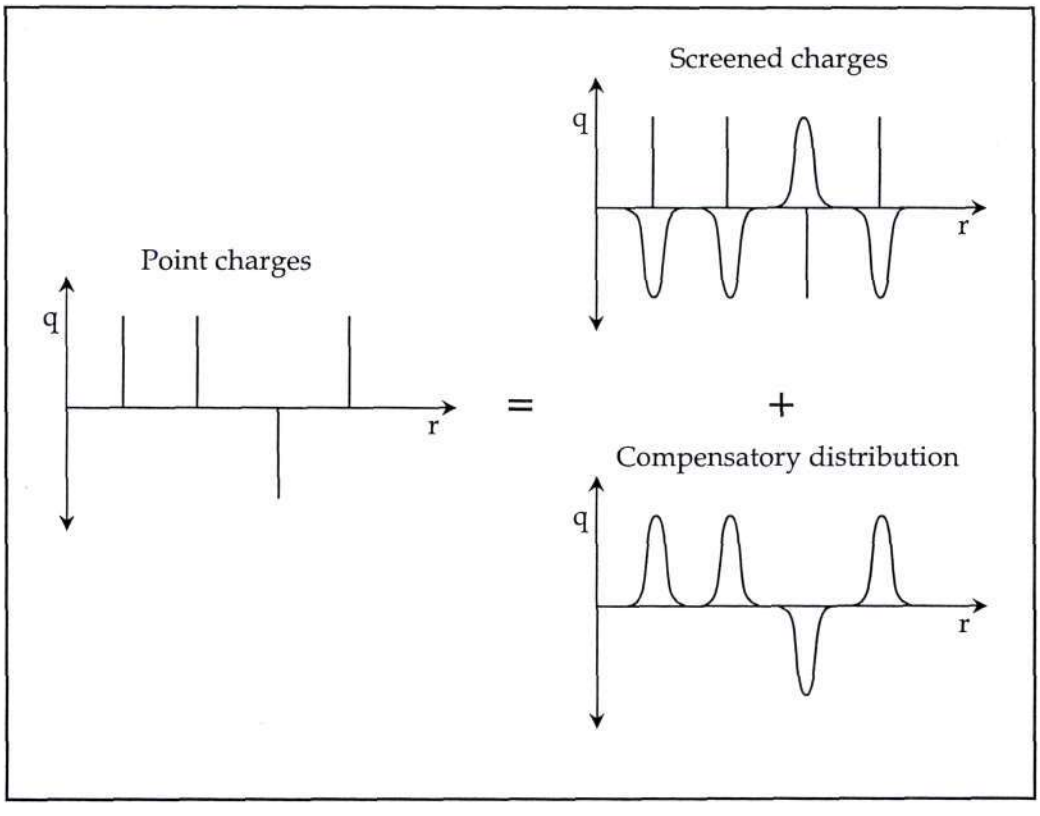

Figure 5.7: Diagram showing the screened point charges and the canceling charge distribution of the Ewald summation method (reproduced from Frenkel and Smit (2002)).

The final form of the summation contains three terms: the contribution of the point charges and the associated screening distribution (a real-space sum), the contribution from the canceling distribution (a reciprocal-space (or Fourier-space) sum) and the correction for the self-interactions discussed above (this is subtracted away). Mathematically, this is written as follows (Allen and Tildesley, 1987):

$$U^{charge} = \frac{1}{8\pi}\sum_{i,j=1}^{N}\sum_{\mathbf{n}=0}^{\infty} q_i q_j \frac{\operatorname{erfc}\left(\kappa\left|\mathbf{r}_{ij}+\mathbf{n}\right|\right)}{\left|\mathbf{r}_{ij}+\mathbf{n}\right|} + \frac{1}{2L^3}\sum_{i,j=1}^{N}\sum_{\mathbf{k}\neq 0}\frac{q_i q_j}{k^2}\exp\left(\frac{-k^2}{4\alpha}\right)\cos\left(\mathbf{k}.\mathbf{r}_{ij}\right) - \frac{\kappa}{8\pi^{3/2}}\sum_{i=1}^{N} q_i^2 \tag{5.28}$$

where erfc (x) is the complementary error function, κ is defined as the square root of α and $\mathbf{k}$ is a set of reciprocal vectors ($\mathbf{k} = 2\pi\mathbf{n}/L^2$). A detailed description and derivation of each term in Equation 5.28 is given in Frenkel and Smit (2002).

The error function falls to zero as x increases and, therefore, if a sufficiently large value for κ is chosen, the only term that will contribute to the real space sum is the term corresponding to $n = 0$ (Allen and Tildesley, 1987). In other words, the first term is reduced to the familiar minimum image convention. However, the sum over the reciprocal vectors (the second term) requires many k-space terms for the sharp charge distribution generated when κ is large (Allen and Tildesley, 1987). Thus, when setting the value of κ for a particular simulation, care must be taken to balance the demands of accuracy with the desire to minimize simulation times. In practice, a compromise is achieved by using between 100 and 200 k-vectors (Allen and Tildesley, 1987) and (in this work) setting κL to 5.6. The value for κL remains constant throughout a simulation to ensure that the sharpness of the Gaussian distribution scales with the box size (L being the simulation box length).

6

The Gibbs Ensemble

The computer simulation techniques discussed thus far are extremely effective when dealing with homogeneous systems. However, simulation of first-order phase transitions (i.e. two distinct, coexisting phases separated by an interface e.g. vapour-liquid equilibrium) is not quite as straightforward. The major stumbling block is that computational expense limits the size of the systems that may be studied. A high percentage of molecules reside in or near the interface for small systems (less than 1000 molecules). Clearly, the large fraction of molecules interacting at the phase boundary biases the results and the bulk properties will not be determined accurately. The Gibbs ensemble technique was developed by Panagiotopoulos (1987) and contains many of the advantages of a direct coexistence simulation, and few of the disadvantages (Frenkel and Smit, 2002). The Gibbs ensemble is currently the method of choice for simulating phase equilibria.

6.1 The Gibbs Ensemble Technique

The Gibbs ensemble technique (Panagiotopoulos, 1987) was originally developed as an alternative means of simulating gas-liquid phase coexistence. The method was named after J. W. Gibbs, who first derived the equations that describe the conditions

necessary for the ensemble. The Gibbs ensemble was later extended to permit simulation of binary (and higher-order multicomponent) systems by Panagiotopoulos et al. (1988). The technique simulates each phase as a separate box, replicated to form an infinite lattice representative of the bulk phase. The boxes have no physical contact and, hence, no interface per se. They are, however, intricately linked through the three variables necessary for phase equilibrium: the system temperature, system pressure and the chemical potential.

For two phases to coexist, the following conditions must be met (Frenkel and Smit, 2002): the temperature of both phases must be equal ($T_I = T_{II} = T$), as must the phase pressures ($P_I = P_{II} = P$) and the chemical potentials of each species ($\mu_I{}^{\alpha} = \mu_{II}{}^{\alpha} = \mu^{\alpha}$). This would seem to indicate that an ideal ensemble for simulation of coexisting phases would be one in which T, P and μ are held constant. However, these variables are all intensive and to obtain an appropriate ensemble, at least one extensive variable (e.g. volume, V) should be fixed (Frenkel and Smit, 2002). The Gibbs ensemble comes very close to achieving the ideal of the constant-μPT ensemble by fixing the difference between chemical potentials instead.

Two forms of the Gibbs ensemble method exist: the *NVT*-Gibbs ensemble is used for pure component simulations, and the *NPT*-Gibbs ensemble is used for simulations with two or more components. In both cases, the number of molecules in the system is fixed. Furthermore, there is an effective temperature range that applies to both versions of the Gibbs ensemble; the method performs best at temperatures lying between 50 and 90 % of the critical temperature. Below 50 % the liquid phase is too dense for an adequate number of particle insertions to occur in a reasonable time, and above 90 % the system becomes unstable (the liquid and vapour phases may switch identity). In both cases, it becomes extremely difficult for the system to achieve equilibrium.

Applications of the Gibbs ensemble technique are reviewed by Panagiotopoulos (1992).

6.2 The *NVT*-Gibbs Ensemble

Originally, Panagiotopoulos (1987) introduced the Gibbs ensemble as a blend of the existing canonical (*NVT*), isothermal-isobaric (*NPT*) and grand canonical (μVT) ensembles (Frenkel and Smit, 2002). However, in Appendix H of Frenkel and Smit

(2002), it is demonstrated that the constant volume Gibbs ensemble is rigorously equivalent (in the thermodynamic limit) to the canonical ensemble. Hence, this form of the Gibbs ensemble is referred to as the *NVT*-Gibbs ensemble.

6.2.1 The Partition Function

As for the classical ensembles presented in Chapter 3, the discussion of the *NVT*-Gibbs ensemble begins with the formulation of the appropriate partition function. Once the partition function has been determined, the density probability function and the acceptance criteria for the Gibbs ensemble trial moves may be derived. For a system of N molecules distributed over two volumes (V_1 and V_2, where the total system volume, V, remains constant and is equal to $V_1 + V_2$) at a temperature T, the partition function is expressed as follows (Frenkel and Smit, 2002):

$$Q_{Gibbs}(N,V,T) \equiv \sum_{N_1=0}^{N} \frac{1}{V\Lambda^{3N} N_1!(N-N_1)!} \int_0^V V_1^{N_1}(V-V_1)^{N-N_1} dV_1 \times \int \exp\left[-\beta U(\mathbf{s}_1^{N_1})\right] d\mathbf{s}_1^{N_1} \int \exp\left[-\beta U(\mathbf{s}_2^{N-N_1})\right] d\mathbf{s}_2^{N-N_1} \tag{6.1}$$

where Λ is the thermal de Broglie wavelength, N_1 is the number of molecules in box 1 ($N = N_1 + N_2$) and $\mathbf{s}^N$ are scaled coordinates (necessary because both box volumes shift as the simulation proceeds) defined as $\mathbf{r}_i = L\mathbf{s}_i$ for i = 1, 2, ..., N by Frenkel and Smit (2002).

It is important to note that for Equation 6.1, the molecules in both boxes are subject to the same intermolecular interactions. According to Frenkel and Smit (2002), the probability of finding a configuration with N_1 molecules in box 1, with a volume V_1 and positions $\mathbf{s}_1^{N1}$ and $\mathbf{s}_2^{N\text{-}N1}$ is:

$$\rho_{pf}(N_1,V_1,\mathbf{s}_1^{N_1},\mathbf{s}_2^{N-N_1}) \propto \frac{V_1^{N_1}(V-V_1)^{N-N_1}}{N_1!(N-N_1)!} \exp\left(-\beta\left[U(\mathbf{s}_1^{N_1}) + U(\mathbf{s}_2^{N-N_1})\right]\right) \tag{6.2}$$

The density probability function given in Equation 6.2 may now be used to derive the acceptance criteria for the Gibbs ensemble trial moves.

6.2.2 Monte Carlo Simulation Scheme

The Gibbs ensemble requires a Monte Carlo scheme that provides satisfactory sampling of all possible configurations of two simulation boxes capable of exchanging molecules and volume. The initial scheme given below applies to a system of particles, which is then extended to molecular systems through implementation of the configurational-bias Monte Carlo (CBMC) method (refer to Section 4.2 for further details on CBMC).

In the Monte Carlo algorithm for the *NVT*-Gibbs ensemble, the following trial moves are considered (Frenkel and Smit, 2002):

1. Displacement of a randomly selected particle.
2. Random volume change (conducted such that $V = V_1 + V_2$ remains constant).
3. Transfer of a randomly selected particle from one box to the other.

During each Monte Carlo cycle, N randomly selected trial moves are attempted. The particle displacements achieve thermal equilibrium and refer to changes of position (i.e. when a particle is transferred to a new location within a particular box) as well as, in the case of molecules, shifts in orientation. Mechanical and chemical equilibrium are attained through the volume and transfer moves, respectively. As before, a move is accepted or rejected based on an acceptance probability derived from the detailed balance condition (detailed balance should always be adhered to).

6.2.3 *NVT*-Gibbs Ensemble Trial Moves

As mentioned in the previous section, the trial moves are based on the condition of detailed balance:

$$\frac{\Xi(o \rightarrow n)}{\Xi(n \rightarrow o)} = \frac{\rho_{pf}(n)}{\rho_{pf}(o)} \tag{6.3}$$

and the derived acceptance criteria are of the form:

$$\Xi(o \rightarrow n) = \min\left(1, \frac{\rho_{pf}(n)}{\rho_{pf}(o)}\right) \tag{6.4}$$

6.2.3.1 *Particle Displacement*

For this particular trial move, the new configuration (n) is obtained from the old configuration (o) via the displacement of a randomly selected particle (see Figure 6.1). For convenience, the particle is taken to be in box 1, but the methodology is equally valid for a box 2 particle (as shown in Figure 6.1).

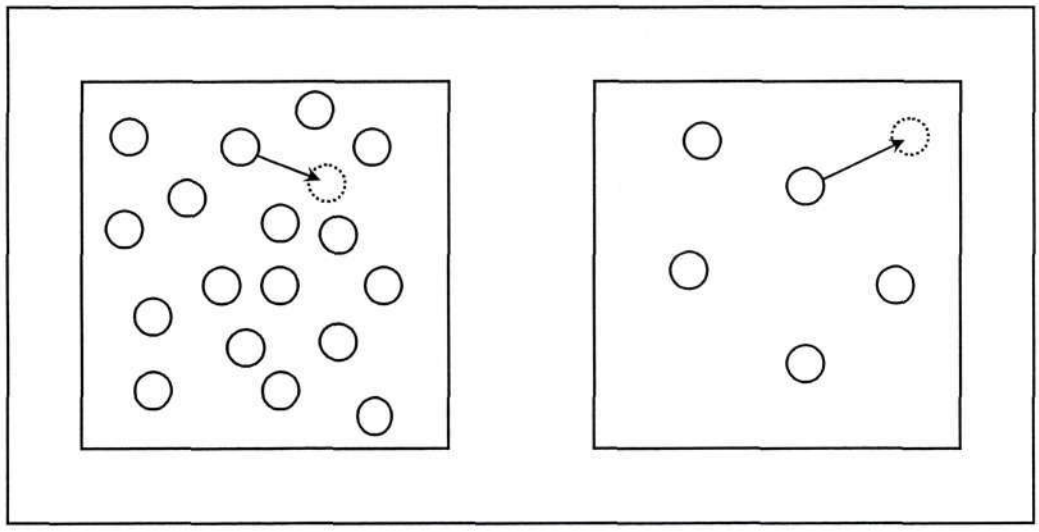

Figure 6.1: Diagrammatic representation of the particle displacement trial move.

The ratio of the probability distributions of states n and o is (Frenkel and Smit, 2002):

$$\frac{\rho_{pf}(n)}{\rho_{pf}(o)} = \frac{\exp[-\beta U(\mathbf{s}_n^{N_1})]}{\exp[-\beta U(\mathbf{s}_o^{N_1})]} \tag{6.5}$$

Substitution of Equation 6.5 into Equation 4.14 yields the acceptance rule:

$$\Xi(o \rightarrow n) = \min\left(1, \exp\{-\beta[U(\mathbf{s}_n^{N_1}) - U(\mathbf{s}_o^{N_1})]\}\right) \tag{6.6}$$

which is identical to that used for particle displacements in the conventional canonical ensemble (Frenkel and Smit, 2002).

6.2.3.2 *Volume Change*

Again, the trial move is derived for box 1, but since the total volume is conserved the move implicitly applies to both boxes (Figure 6.2).

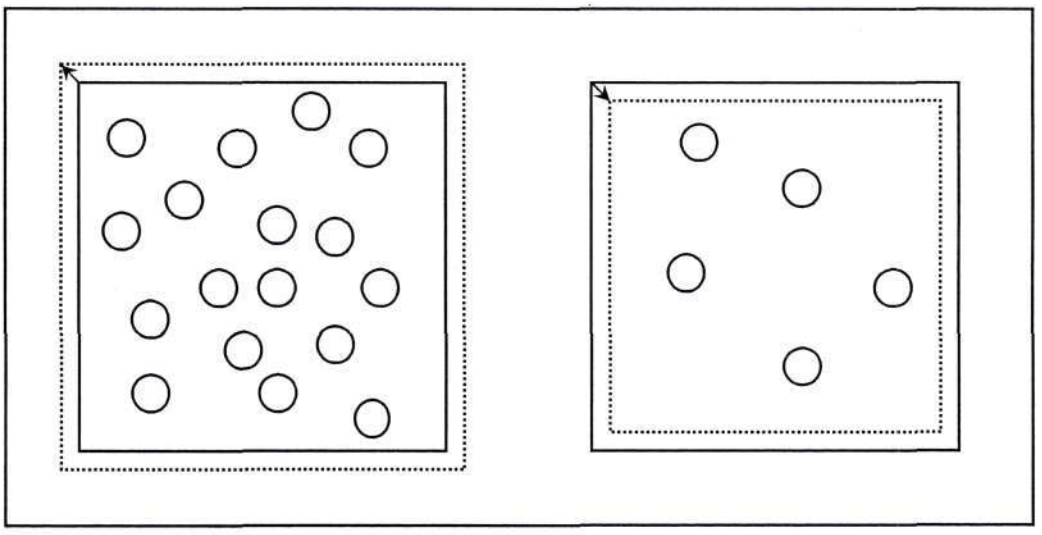

Figure 6.2: Diagrammatic representation of the volume change trial move.

Consider a volume change, ΔV, such that $V_1^n = V_1^o + \Delta V$. The probability ratio is (Frenkel and Smit, 2002):

$$\frac{\rho_{pf}(n)}{\rho_{pf}(o)} = \frac{\left(V_1^n\right)^{N_1}\left(V - V_1^n\right)^{N-N_1} \exp[-\beta U(\mathbf{s}_n^N)]}{\left(V_1^o\right)^{N_1}\left(V - V_1^o\right)^{N-N_1} \exp[-\beta U(\mathbf{s}_o^N)]} \tag{6.7}$$

and the associated acceptance criterion for the volume change trial move is:

$$\Xi(o \to n) = \min\left(1, \frac{\left(V_1^n\right)^{N_1}\left(V - V_1^n\right)^{N-N_1}}{\left(V_1^o\right)^{N_1}\left(V - V_1^o\right)^{N-N_1}} \exp\{-\beta[U(\mathbf{s}_n^N) - U(\mathbf{s}_o^N)]\}\right) \tag{6.8}$$

Equations 6.7 and 6.8 apply to the volume change move originally proposed by Panagiotopoulos (1987). Frenkel and Smit (2002) suggest that a more natural choice for creating the new configuration in the volume change trial move is to generate a random walk in $\ln[V_1/(V - V_1)]$ rather than $\ln V_1$. The advantages offered by this alternative volume move are that the domain of this random walk coincides with all values of V_1 and, in addition, the average step size is less sensitive to the density (Frenkel and Smit, 2002). The modified probability function for a configuration n with volume V_1 is now written as:

$$\rho_{pf}(n) \propto \frac{\left(V_1^n\right)^{N_1+1}\left(V - V_1^n\right)^{N-N_1+1}}{V N_1!(N - N_1)!} \exp[-\beta U(\mathbf{s}_n^N)] \tag{6.9}$$

and by imposing detailed balance, the following acceptance rule arises:

$$\Xi(o \to n) = \min\left(1, \left(\frac{V_1^n}{V_1^o}\right)^{N_1+1}\left(\frac{V - V_1^n}{V - V_1^o}\right)^{N-N_1+1} \exp\{-\beta[U(\mathbf{s}_n^N) - U(\mathbf{s}_o^N)]\}\right) \tag{6.10}$$

It should be noted that the modified volume change trial move does not effect either the particle displacement or particle exchange moves (Frenkel and Smit, 2002).

6.2.3.3 Particle exchange

The particle exchange trial move occurs when a particle is transferred from one box to the other (Figure 6.3). The derivation given below is for the case where a particle is removed from box 1 and inserted into box 2 and the corresponding density probability function is (Frenkel and Smit, 2002):

$$\frac{\rho_{pf}(n)}{\rho_{pf}(o)} = \frac{N_1!(N - N_1)! V_1^{N_1-1}(V - V_1)^{N-(N_1-1)} \exp[-\beta U(\mathbf{s}_n^N)]}{(N_1 - 1)!(N - (N_1 - 1))! V_1^{N_1}(V - V_1)^{N-N_1} \exp[-\beta U(\mathbf{s}_o^N)]} \tag{6.11}$$

The associated acceptance rule takes the following form:

$$\Xi(o \rightarrow n) = \min\left(1, \frac{N_1(V - V_1)}{(N - N_1 + 1)V_1} \exp\{-\beta[U(\mathbf{s}_n^N) - U(\mathbf{s}_o^N)]\}\right) \tag{6.12}$$

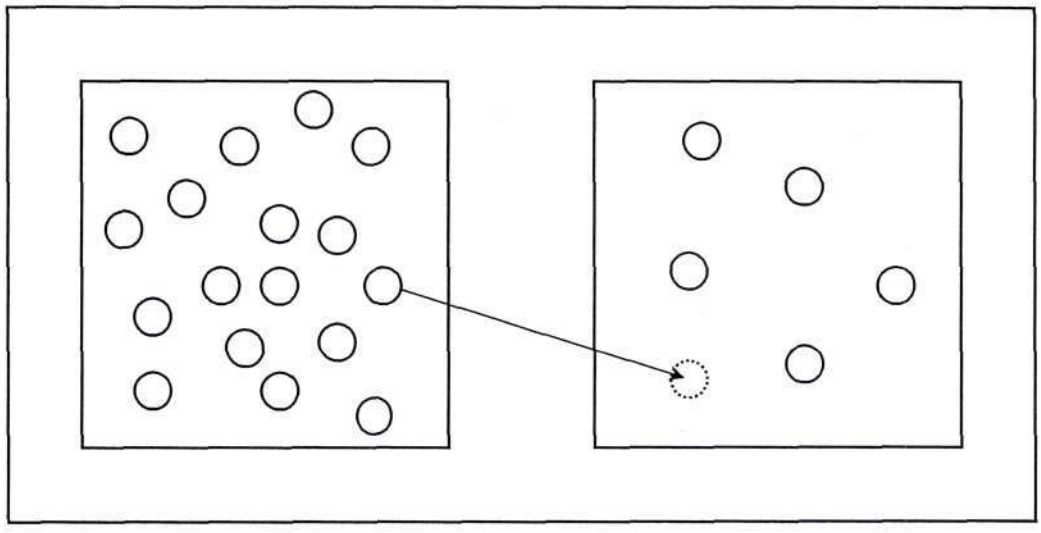

Figure 6.3: Diagrammatic representation of the particle exchange trial move.

The particle (or molecule) exchange move is an integral facet of the Gibbs ensemble and it is crucial that a sufficient number of so-called swap moves occur during a simulation. A generally accepted figure (Martin and Siepmann, 1998) is one molecule exchange move for every ten Monte Carlo cycles (on average). Clearly, the greater the structural complexity of the molecules in the simulation, the more difficult it becomes to achieve a satisfactory molecular transfer acceptance rate. The use of the CBMC technique greatly improves the probability that an exchange move will be successful.

6.2.4 CBMC in the *NVT*-Gibbs Ensemble

If the *NVT*-Gibbs ensemble is implemented without the use of CBMC, the particle exchange (or swap) move is accepted extremely seldom for systems involving long-chain molecules. For this reason, application of the Gibbs ensemble was initially

restricted only to systems containing atoms or small molecules (Frenkel and Smit, 2002). The introduction of the CBMC technique, however, permitted simulation of more complex molecules using the Gibbs ensemble.

The particles (or simple molecules) inserted in the exchange move described above (Section 6.2.3.3) were placed randomly. Using CBMC, the molecules are now regrown atom by atom in a randomly selected box (Frenkel and Smit, 2002). Assuming this box to be box 1 with volume V_1 and containing N_1 molecules, the algorithm proceeds (Frenkel and Smit, 2002) in a similar manner to the procedure outlined in Section 4.2.2:

1. The first bead or atom is inserted at a random position, and the external energy and associated Rosenbluth weight of this new configuration are calculated:

$$w_1^{ext} = k \exp[-\beta U_1^{ext}(n)] \tag{6.13}$$

2. The next atom, i, is inserted by generating k trial orientations based on a probability that is a function of the bonded intramolecular energy (Section 5.2.1.1):

$$\mathrm{P}_i^{bond}(b_n) = C \exp[-\beta U_i^{bond}(b_n)] \tag{6.14}$$

The external energy is calculated for each trial, and the summed factor:

$$w_i^{ext}(n) = \sum_{j=1}^{k} \exp[-\beta U_i^{ext}(b_j)] \tag{6.15}$$

and one of the k trial orientations is selected based on the probability:

$$\mathrm{P}_i^{ext}(b_n) = \frac{\exp[-\beta U_i^{ext}(b_i)]}{w_i^{ext}(n)} \tag{6.16}$$

3. Step 2 is repeated until the entire molecule has been regrown. The Rosenbluth factor of the whole molecule may now be determined by taking the product of the Rosenbluth weights corresponding to the individual atoms:

$$W^{ext}(n) = \prod_{i=1}^{l} w_i^{ext}(n) \tag{6.17}$$

For the other box, the Rosenbluth factor of a randomly selected molecule must be determined. This is achieved via the following procedure:

1. A molecule is randomly selected.

2. The external energy and Rosenbluth weight of the first atom are calculated:

$$w_1^{ext}(o) = k \exp[-\beta U_1^{ext}(o)] \tag{6.18}$$

3. For the next atom, the actual position (b_o) and the k - 1 trial orientations (generated with a probability given by Equation 6.14) form a set of trial configurations, $\{b'\}_k$, that are used to determine the factor:

$$w_i^{ext}(o) = \exp[-\beta U_i^{ext}(o)] + \sum_{j=2}^{k} \exp[-\beta U_i^{ext}(b_j')] \tag{6.19}$$

4. Step 3 is now repeated until the entire molecule chain has been traced and its Rosenbluth factor found:

$$W^{ext}(o) = \prod_{i=1}^{l} w_i^{ext}(o) \tag{6.20}$$

The move is then accepted with probability:

$$\Xi(o \to n) = \min\left(1, \frac{V_1(N - N_1)}{(V - V_1)(N_1 + 1)} \frac{W^{ext}(n)}{W^{ext}(o)}\right) \tag{6.21}$$

For proof of the validity of this algorithm, the following manuscripts may be consulted: Mooij et al. (1992) and Smit et al. (1995, 1998).

6.2.5 Chemical Potential in the Gibbs Ensemble

One of the Gibbs ensemble trial moves involves insertion of a molecule into a particular phase. To determine whether the move is accepted or not, the energy associated with the insertion must be calculated. This energy is equivalent to that of a test (also known as a ghost) molecule, and may, therefore, be used in the Widom insertion method (Widom, 1963) to calculate the chemical potential without incurring additional computational expense (Frenkel and Smit, 2002). It is not necessary to determine the chemical potential in the Gibbs ensemble method, but it is a useful property to calculate. It is important to check that the chemical potentials of the

individual phases are being computed correctly and, in addition, the chemical potential serves as a valuable means of testing whether the system has reached equilibrium or not. An ensemble average for the chemical potential of box 1 may be determined using the following expression (Frenkel and Smit, 2002):

$$\mu_1 = -k_B T \ln \frac{1}{\Lambda^3} \left\langle \frac{V_1}{N_1 + 1} \exp[-\beta \Delta U_1^+] \right\rangle_{Gibbs,box1} \tag{6.22}$$

where ΔU_1^+ is the energy of a test (or ghost) molecule in box 1 and $<\ldots>_{Gibbs,box1}$ denotes a box 1 ensemble average in the Gibbs ensemble. This definition applies for systems in which the two boxes do not change identity once equilibrium has been achieved (i.e. at temperatures sufficiently far below the critical temperature). Smit and Frenkel (1989) discuss the more general case in which the box identities may shift.

It should be noted that the original Widom expression is only valid in a constant-*NVT* ensemble, but may be modified for application to other ensembles (Frenkel and Smit, 2002).

6.3 The *NPT*-Gibbs Ensemble

At the start of the chapter, the *NPT*-Gibbs ensemble was introduced as the multicomponent form of the Gibbs ensemble. The reason the constant-*P* form of the ensemble applies only to systems containing two or more components is simple: degrees of freedom. A pure component, two-phase system has only one degree of freedom. Thus, if the temperature is specified, then the pressure obviously cannot be. In other words, for a pure component system, the two-phase region is a line in the *P-T* plane. Consequently, the probability that a specific choice for both temperature and pressure will lie on this line (i.e. at the phase transition) is extremely remote (Frenkel and Smit, 2002). However, the extra degree of freedom afforded by a two-component (or higher) system relaxes this restriction, and both pressure and temperature may now be specified. Continuing the geometric description of this problem, for a binary system the two-phase region corresponds to a finite area in the *P-T* plane (instead of a line).

Clearly, for the *NPT*-Gibbs ensemble the total system volume is no longer conserved and the box volumes are permitted to fluctuate independently of one another

(Panagiotopoulos et al., 1988). This requires a slight modification to the acceptance criterion for the volume change trial move. The remaining trial moves are not affected in any way.

As before, the trial move is considered to occur in box 1 for convenience, but could equally well occur in box 2 or even in both boxes simultaneously. However, Panagiotopoulos et al. (1988) note that the overall convergence will occur more rapidly if the box volumes are changed autonomously (i.e. one at a time). Consider a volume change ΔV in V_1 such that $V_1^n = V_1^o + \Delta V$. The corresponding acceptance rule for this trial move is (Panagiotopoulos et al., 1988):

$$\Xi(o \rightarrow n) = \min\left(1, \left(\frac{V_1^n}{V_1^o}\right)^{N_1} \exp\left\{-\beta[U(s_n^{N_1}) - U(s_o^{N_1}) + P\Delta V]\right\}\right) \tag{6.23}$$

7

Beowulf Clusters

Named after a Scandinavian hero immortalized in a poem recounting his many exploits (the most famous of which was the slaying of the monster Grendel), the first Beowulf cluster was built in 1994. This initial cluster was constructed primarily to demonstrate that off-the-shelf computer hardware could be used to tackle serious computational problems, rather than relying on the supercomputers that had been used for these calculations up to that point. The rapid increase in computing power over the last decade has made Beowulf computers more and more applicable for high-end tasks such as the study of atmospheric chemistry and molecular simulation. All of the simulation work conducted in this project was performed on a Beowulf cluster named Yoda, situated in the laboratories of the School of Chemical Engineering at the University of KwaZulu-Natal, Durban.

7.1 Beowulf Cluster History

The original concept for creating a computer cluster of equivalent computing power to that of extremely costly supercomputers was formulated by Donald Becker at NASA in the summer of 1994 (Duggan, 2001). In conjunction with Thomas Sterling, Becker created a parallel processing computer consisting of sixteen 486 DX4 processors (linked

via a novel channel bonding method allowing them to tie up multiple 10 MBit/s Ethernets without additions of expensive switching systems (Askåsen and Johansen, 2002)) that was capable of performing rigorous computations. They decided to name their cluster Beowulf (the name has been used ever since to describe this genre of computer cluster) and used it to model the fragmentation of the Shoemaker-Levy 9 comet as it passed Jupiter and was pulled apart by the massive gravitational field of the planet (Duggan, 2001).

An idea of the computational power that may be achieved using Beowulf clusters is provided by the ASCI Red TFLOPS supercomputer that was put together in a joint effort between Intel and the American DoE. First run in 1996, the computer was a Beowulf-type cluster consisting of 9 216 Pentium Pro processors (Duggan, 2001) and reached the incredible computational speed of 1×10^{12} floating point operations per second (that is, 1 Teraflop/s).

7.2 Beowulf Cluster Architecture

Beowulf clusters are divided into two broad categories, referred to as Class I and Class II (Radajewski and Eadline, 1998). Class I clusters are characterized by the fact that they are constructed from commodity computer components. The hardware generally consists of computer processors (either Intel Pentium or AMD), conventional EIDE hard drives and the associated motherboards and RAM (Random Access Memory) necessary to run the node in conjunction with the chosen processor chipset, all of which may be purchased from any number of computer hardware suppliers. All of these components are inexpensive because they are mass-marketed and produced in sufficiently large volumes. Some Beowulf clusters use symmetric multi-processor nodes which provide an incredible boost in performance for the slight increase in price, according to Askåsen and Johansen (2002). Class II Beowulf clusters differ in that certain components are non-commodity items. This provides a significant increase in performance, but at the same time there are consequences (Radajewski and Eadline, 1998): system cost rises, dependence on a single hardware vendor, driver support may vary and maintenance of the cluster becomes more difficult.

All Beowulf clusters display similar system architecture (Figure 7.1). The cluster center is the master node, on which the majority of the software essential for the

running of the cluster is installed. The master node is responsible for distributing the jobs that are run on the cluster and controlling the functioning of the cluster as a whole. The actual computations take place on the client nodes, which vary greatly in number depending on the cluster application and the available budget. The master node communicates with the client nodes via a network switch using a message passing interface (MPI), which clearly implies that every node must contain an Ethernet card (usually a 100 Mbit/s card). A second network card is usually installed in the master node to allow access to the outside world and simultaneously permitting remote cluster access from a desktop workstation. Such access is password protected and is normally controlled using Secure Shell (SSH) software.

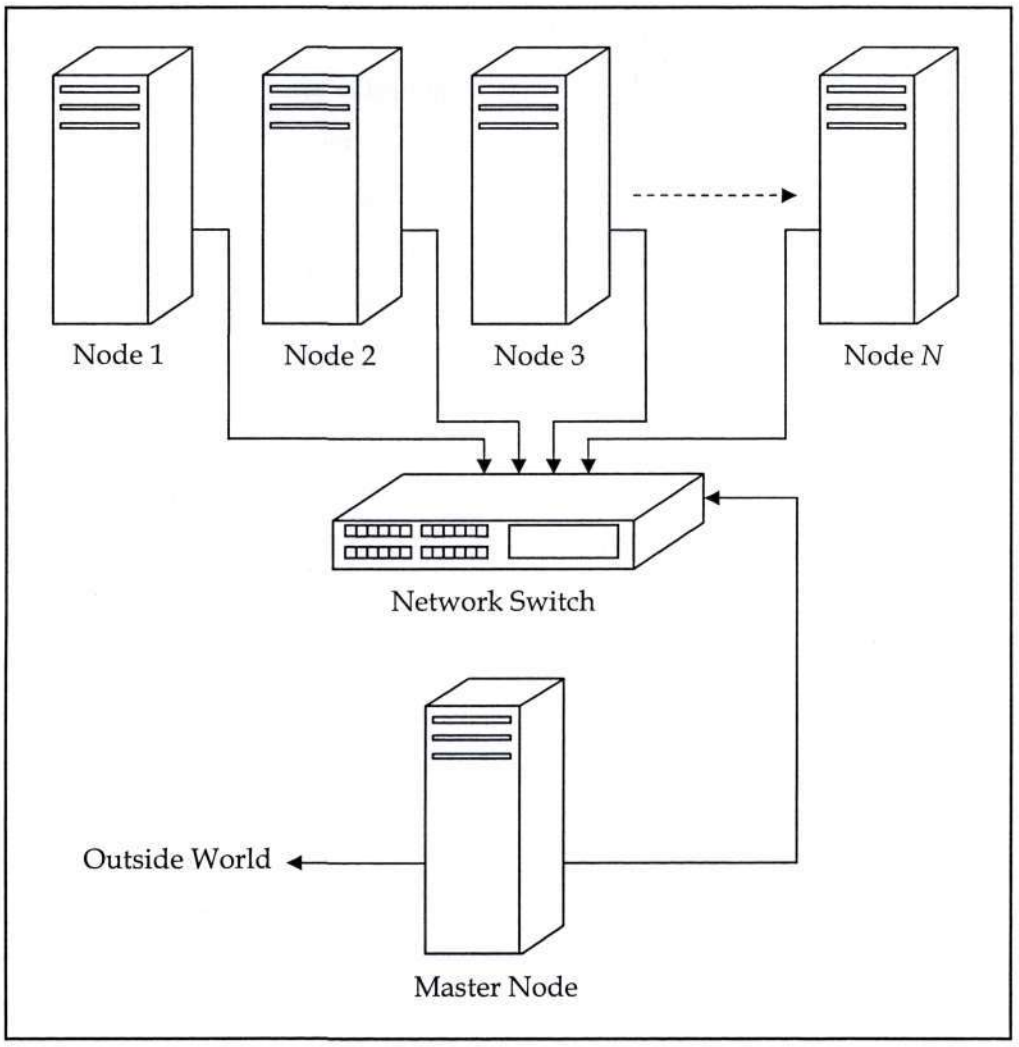

Figure 7.1: Schematic diagram showing the general structure of a Beowulf cluster.

7.2.1 The Operating System

Typically, Beowulf clusters are run using one of the many Linux versions that are available. The Linux operating system was initially written by Linus Torvalds in the early nineties, with version 1.0 released in 1994 (Askåsen and Johansen, 2002). The Linux code has always been open, meaning that the source code of the operating system may be freely downloaded, modified and distributed. It is this freedom that has led to the many versions of Linux that are currently available, providing the flexibility that is one of the main strengths of the operating system (and, incidentally, the reason why the code is relatively bug-free). This does not, of course, mean that all Linux distributions are free - some must be purchased. However, the source code is always accessible once the operating system has been bought. Linux is developed and distributed under the GNU General Public Licence (Askåsen and Johansen, 2002) and has proved popular for a tremendous variety of computer applications.

The Linux operating system that appears to be used most often for the purposes of Beowulf clusters and their fairly specific networking requirements is the Red Hat distribution. The latest version of Red Hat (version 9.2) may be downloaded off the internet (go to `www.redhat.com` then follow the link to the Fedora Project site), free of charge. Later versions of Red Hat than 9.2 do exist, but they are no longer named Red Hat x.x, but Fedora Core instead.

7.2.2 The Message Passing Interface

The message passing interface (MPI) is used, as was mentioned above, primarily to allow effective communication between the various nodes making up the cluster. Quick, efficient message passing is essential for running computations in parallel (i.e. dividing a calculation into several parts, which are then executed on separate nodes before being recombined). This ability to run parallel jobs is one of the main reasons why Beowulf clusters are able to compete computationally with conventional supercomputers. A message passing interface refers specifically to a message passing library specification that is freely available for implementation by hardware and software vendors (Askåsen and Johansen, 2002). According to Askåsen and Johansen (2002), three fundamental functions of MPI are worth noting:

1. MPI takes care of the message passing requirements of parallel computations, permitting synchronization and data exchange between processes running on physically separate client nodes,
2. MPI is a widely transferable collection of well-defined subroutines, i.e. the interface can be successfully installed on various systems and platforms without the underlying software requiring major alteration or modification,
3. The MPI specification is definitely not constrained to a particular system (or platform) and specific programming language (MPI routines are available for many of the conventional languages such as C++ and Fortran).

More details regarding the MPI standard and other MPI documents may be downloaded free of charge from the MPI website: `http://www.mpi-forum.org`. An alternative to MPI that is also commonly used for communication in Beowulf clusters is known as Parallel Virtual Machine (PVM). MPI is focused on here because Yoda used MPI rather than PVM for message passing.

7.2.3 Advantages and Uses of Beowulf Clusters

The main advantage of Beowulf clusters over traditional supercomputers is their price: they are comparatively very cheap. Moreover, due to the constantly increasing power of commodity computing hardware, this saving in terms of capital outlay does not have a corresponding negative impact on performance. In fact, Radajewski and Eadline (1998) estimate that the price to performance ratio of Beowulf clusters compared with conventional supercomputers is three to ten times better. It is also fairly straightforward to add nodes to a cluster (or replace old nodes with new, up-to-date computers), thereby improving the overall capability of the cluster. Another major positive is that all of the software necessary to effectively operate a Beowulf cluster is open-source and can be downloaded free of charge (provided, usually, that it is to be used for non-commercial purposes).

The areas in which Beowulf clusters find application are many and varied. From the example given in Section 7.1, in which the first ever Beowulf cluster was used to model the Shoemaker-Levy 9 comet as it fragmented, to investment (stock) trend analysis and medical research. Furthermore, the possibilities offered by clusters such as Yoda are encouraging more and more academic institutions to develop Beowulf clusters for

research purposes. Certainly, the modeling and simulation of physical and chemical phenomena that are difficult to study using traditional methods may be effectively, efficiently and safely investigated using Beowulf clusters and the techniques of molecular dynamics and molecular simulation.

7.3 Yoda Beowulf Cluster

The Yoda Beowulf cluster developed at the School of Chemical Engineering at the University of KwaZulu-Natal (UKZN) originally consisted of twenty computers. Yoda followed the customary Beowulf cluster architecture, with one of the twenty computers selected to serve as the master node and the other nineteen installed as client nodes. During this project, it became apparent that an upgrade was necessary and ten new nodes were purchased and integrated into the existing cluster.

7.3.1 Original Hardware

The hardware specifications for the twenty computers that were purchased initially are outlined in Table 7.1 below:

Table 7.1: Hardware specifications for the original nodes forming Yoda.

Component	Specification
Processor (CPU)	AMD Thunderbird 1.2 GHz
Motherboard	Asus A7V-E
RAM	256 MB DIMM SDRAM
Hard Drive	20 GB
Network Card	3Com Etherlink (10/100 Mbit/s)

The master node differed slightly from the client nodes in that it contained an 80 gigabyte (GB) hard drive. The increased disk space was necessary because all of the simulation output files (in addition to the Linux system files and other installed software) were stored on the master node. The master node tower also contained a CD-ROM for software installation and essential system maintenance. A second 10/100 MBit/s network card (D-Link DFE-570TX) was installed to allow the master node to connect to the UKZN LAN (Local Area Network). Furthermore, this permitted remote access from an office-based desktop workstation using the Secure Shell (SSH) software (discussed further, below). A dedicated monitor, keyboard and mouse were attached to the master node to allow ready access to Yoda for routine maintenance.

The master node communicated with the slave nodes using a 24 port 3Com 3300 XM 10/100 Mbit/s network switch. All of the cluster connections were made using ready-made, industrially crimped CAT.5e cables to reduce network lag induced through message passing to an absolute minimum.

7.3.2 Software and Applications

Initially, the Scyld Beowulf Scalable Distribution (a modified version of Linux Red Hat 6.2) was installed on Yoda. The operating system was later updated to the NPACI Rocks Cluster Distribution (Matterhorn 3.1.0). As part of the Rocks installation a number of software rolls were loaded onto the master node, viz. Rocks base, HPC, SGE, Java and the Intel compiler rolls. A number of MPI versions existed on the cluster, but the MPICH version with GNU compilers was the message passing interface of choice.

MPI is required whenever simulations are run in parallel. However, the simulations performed during this project were all serial. In most cases, Beowulf clusters are used for parallelization; speeding up computation times for large, complex problems with thousands of molecules that would otherwise take weeks to complete. In contrast, the simulations performed in this study were relatively short (the longest taking approximately four days) with the maximum number of molecules not exceeding 800 (see Table 8.12). Furthermore, it was found that running a job under these conditions (i.e. small system sizes) achieved an increase in speed of only 1.7 using two nodes in parallel. Obviously, it is a more efficient use of system resources to simply run separate

jobs co-currently on these two nodes, provided the simulations are comparatively short. Hence, serial simulations were used to generate the data presented in Chapter 8. These serial jobs were submitted and distributed to the client nodes using the Sun Grid Engine 5.3 application.

As has been mentioned previously, remote access to the cluster via the UKZN LAN was achieved using the Secure Shell (SSH) software. SSH allows the cluster to be accessed in two different manners: first, a remote shell can be used to submit jobs and for general communication with the master node and second, a secure file transfer protocol (sftp) client window can be opened, permitting files to be copied on to and off of Yoda (Askåsen and Johansen, 2002). The main advantage of using SSH is that it allows the Linux environment of the cluster to be accessed from the Windows-based desktop workstation environment. Both SSH applications are tightly password-protected. The combined SSH client bundle can be downloaded free of charge (for non-commercial purposes) from `http://www.ssh.com`.

The molecular simulation code used to produce the data presented in this thesis was the MCCCS Towhee code, written in conjunction with several researchers from the well-known Siepmann research group and currently maintained and enhanced by Marcus Martin (Martin et al., 2005). New versions of Towhee are constantly being produced as the source code is improved, bugs are removed and additional ensembles and potential models are programmed in. The Towhee source code, associated applications, the various user manuals and details regarding the code and its capabilities may be found at `http://towhee.sourceforge.net/`. Details on how to compile and install MCCCS Towhee are also given on the site.

7.3.3 Cluster Upgrade

One of the advantages of Beowulf clusters discussed in Section 7.2.3 was the relative ease with which new computers may be added to the cluster. This project was focused on the molecular simulation of polar compounds, however, the charged interactions of these species result in much longer simulation times (shown in Table 8.12) than for non-polar compounds like alkanes. Consequently, it was decided to purchase ten new computers and integrate them into the existing cluster in order to improve the simulation times. Due to a lack of available space the new nodes were not simply

added on, but rather replaced ten of the existing nodes. The specifications of the new nodes are given in Table 7.2:

Table 7.2: Hardware specifications for the new nodes integrated into Yoda.

Component	Specification
Processor (CPU)	AMD Athlon 2.8+ GHz (64 bit)
Motherboard	ASRock K8Upgrade-760GX
RAM	512 MB DDR400
Hard Drive	40 GB
Network Card	10/100 Mbit/s Ethernet (Onboard)

The process of adding new nodes to the cluster is relatively straightforward using the Rocks Distribution. However, the fact that several nodes were going to be removed first increased the difficulty somewhat. To avoid complications due to computer name duplication, all of the nodes were removed from the cluster and the new nodes (and the old nodes that were to be kept) were added from scratch. To carry out any of these operations, the user must first login as `root`. Thereafter, the `insert-ethers` command is used to first remove and then add the client nodes. To remove a node, the `insert-ethers --replace="compute-0-x"` command is executed (where `x` is the appropriate numeral for the node being removed). At the on-screen prompt which then appears, press enter, followed by *F1*. This will remove the particular client node from the MySQL database on the master node (also sometimes referred to as the frontend). To update the relevant configuration files (such as `/etc/hosts` and `/etc/dhcpd.conf`), the files need to be rebuilt by executing the following command: `insert-ethers --update`. This procedure (taken from the NPACI Rocks website: `http://rocks.npaci.edu/papers/rocks-documentation`) is repeated until all of the client nodes have been removed.

The appropriate nodes are then added to the cluster, following the procedure recommended in the NPACI Rocks user guide (2002). Briefly, the `insert-ethers` command is executed from the master node and the default selection, `Compute`, is

selected. The first node to be added is then switched on and the programme automatically captures the client node DHCP requests and stores this information in the Rocks MySQL database before updating the necessary system files. The nodes are also automatically named (`compute-0-x`, with `x` running from 0 to *N*) in the order that they are switched on. The new and old nodes were differentiated by installing them into two separate cabinets by executing `insert-ethers --cabinet=1` when adding the old nodes, which then received names in the form `compute-1-x`. The NPACI Rocks Distribution kernel was installed on the client nodes across the network using PXE (i.e. network) booting. Although this option makes maintenance of Yoda easier, some complications did arise for network cards with later versions of PXE (2.20 and higher). This problem was overcome by commenting out lines 145 and 146 of the `/opt/rocks/lib/python/rocks/reports/dhcpd.py` file and restarting the DHCP daemon.

Figure 7.2: Photo of the Beowulf cluster, Yoda, showing the master node, client nodes, network switch and UPS units.

During the course of this project and the required routine administration and maintenance of Yoda, it was found that the main source of problems was due to brief power dips or spikes that would cause the computers to reboot (often while running simulations). Normally this would have no effect aside from the irritation of having jobs interrupted; however, as the computers became older, these power fluctuations would often cause hardware failure (usually in the form of hard drive corruption). This was another reason why the new nodes were purchased. In addition, four UPS units (each capable of delivering power to four computers) were bought and installed. These line-interactive UPS units smoothed out the power supply and reduced the frequency with which the cluster rebooted tremendously.

A picture of Yoda is shown in Figure 7.2 on the facing page. The new and old nodes are clearly visible (the black/silver and plain white towers, respectively), as are the UPS units (top left), the master node (top right) and the network switch (between the UPS units and the master node).

7.3.4 Common Problems

The most common problems with Yoda encountered during this project were all hardware related. Computer components have a limited lifespan, and it became apparent that the old node hardware was beginning to fail more and more regularly. As was mentioned in the preceding section, this fact urgently emphasized the need for new client nodes to be purchased.

In general, the hard drives were the most susceptible to failure and frequently became corrupted and unusable if the cluster was switched off without following the correct shutdown procedure (by, for example, a loss of electrical power). However, by the end of the project the RAM, network cards and even motherboards of the old client nodes were starting to fail, sometimes without an obvious cause.

In most cases, it is not too serious if a client node hard drive becomes corrupt, since it can simply be replace and the Rocks Distribution re-installed across the network. However, the master node hard drive eventually became severely corrupted, to the point that the drive could not be backed up due to bad sectors throughout the disk. Careful and painstaking repairs using the `fsck` command (this is a Linux command

designed for checking Linux filesystems for consistency and repairing them if possible: Welsh et al., 2003) were necessary before the hard drive could be effectively backed up.

One particular difficulty that cropped up after the new nodes had been purchased and installed deserves mention. The problem was eventually discovered to be the result of an input/output (i.e. network communication) error caused by a certain node. This error prevented the master node from communicating with this node and on each occurrence an error message was printed to a file in `/etc/var`. Since this was taking place every couple of seconds, the file quickly became extremely large and in the end filled the `hda1` partition on the master node. Naturally, this generated a whole host of related errors, without the true source of the problem being evident. After a great deal of troubleshooting, the input/output error was discovered and corrected; with the problem solution was completed by deleting the huge message file occupying `hda1`.

8

Results and Discussion

This chapter presents the results of the project, and the discussion thereof. The pertinent theory and the motivation for the selection of the pure component and binary systems studied have been discussed in preceding chapters. It has already been noted that this project was divided into two parts. The major focus was molecular simulation of polar compounds, but some modeling of binary carboxylic acid VLE data was also performed. This was a follow-on investigation that stemmed from work conducted during the Masters project that led into the current study. As such, the modeling component of the project will be discussed first.

Although modeling of charged species is important work, the key motivation for this PhD study was to investigate the ability of current molecular simulation techniques to predict phase equilibrium data for these compounds. Hence, the project is heavily weighted toward the molecular simulation aspect of the work. The simulation results will be presented as follows:

- Pure component data
- Parameterization of H_2S, SO_2, NO_2
- Binary VLE data
- Combining rules analysis
- Dimerization of carboxylic acids

In general, the results given here will be plotted rather than tabulated, since a graphical representation is usually more meaningful than a table consisting of only numerical values.

8.1 Chemical Theory Modeling

As was mentioned above, the modeling component of this project was a continuation of work begun during the preceding Masters project (Clifford, 2004). Initially, standard Gibbs excess energy models such as the Wilson (Wilson, 1964), UNIQUAC (Abrams and Prausnitz, 1975) and NRTL (Renon and Prausnitz, 1968) models (with vapour phase imperfections accounted for by the Pitzer-Curl (Pitzer and Curl, 1954) or Hayden and O'Connell (Hayden and O'Connell, 1975) correlations) were used in an effort to accurately model the measured carboxylic acid binary systems, viz. propanoic acid + pentanoic acid and 2-methyl propanoic acid + pentanoic acid. Although these local composition-based, liquid-phase activity coefficient models proved adequate, the fundamental, underlying theory was not strictly applicable to carboxylic acid systems that involve extensive dimerization in both phases. This was particularly true for the vapour phase, since techniques such as that of Hayden and O'Connell do not explicitly take dimers into consideration.

In order to account for the formation of the carboxylic acid dimers and heterodimers, the experimental liquid-phase activity coefficients were determined using a procedure based on the chemical theory for the vapour phase, following on from similar work carried out by Tamir and Wisniak (1975). The experimental activity coefficients, γ_A and γ_B, were then correlated using the commonly utilized Gibbs excess energy models. This proved to be difficult, with the best results being obtained from the T-K Wilson equation (Tsuboka and Katayama, 1975). Nonetheless, it allowed the area test for thermodynamic consistency (strongly dependent on the chemical equilibrium constants, K) to be applied.

To allow the carboxylic acid systems to be modeled, a novel and efficient bubble-point calculation routine incorporating chemical theory was developed. This algorithm yields system pressure and vapour-phase compositions (at any arbitrary liquid composition and temperature) of sufficient accuracy for distillation equipment design.

Details regarding chemical theory and the actual calculation procedure followed here may be found in Chapter 2.

Throughout the results section dealing with the chemical theory modeling of the carboxylic acids, *A* and *B* refer to the monomers of components 1 and 2, respectively.

8.1.1 VLE Data and Activity Coefficients

The VLE data measured for the propanoic acid + pentanoic acid and 2-methyl propanoic acid + pentanoic acid systems (Clifford et al., 2005) are presented in Tables 8.1 to 8.8 and plotted in Figures 8.1 to 8.8. Isobaric VLE data for the test system cyclohexane + ethanol, when compared with the published data from Joseph et al. (2001) and Morachevsky and Zharov (1963), showed excellent agreement and confirmed satisfactory operation of the experimental equipment for highly non-ideal systems.

Included in Tables 8.1 to 8.8 are the species vapour-phase concentrations η_{A_1} and η_{B_1}, obtained by simultaneously solving Equations 2.12 and 2.13, and the liquid-phase activity coefficients γ_A and γ_B obtained from Equation 2.19. Extensive calculations for all systems showed that the use of Equation 2.20 for the heterodimerization constant in the vapour phase produced physically meaningless liquid-phase activity coefficients. Use of the geometric mean (Equation 2.21) proposed by Campbell (1994) for alcohol systems exhibiting extensive dimerization, produced generally satisfactory activity coefficients and, hence, was adopted for all calculations. It was also found that in the very dilute regions the activity coefficients were extremely sensitive to even slight deviations in the liquid-phase compositions. Thus, a few experimental points not lying on a smooth polynomial through the origin gave meaningless activity coefficients. The corrected liquid-phase compositions are denoted by an asterisk in the tables. The marked sensitivity to these very small variations in liquid composition in the dilute regions was unexpected and suggests the need for very careful experimentation in these areas.

Computational experiments with different *K* values from the literature (see Table 2.1) showed slightly superior results for the Design Institute for Physical Properties (DIPPR) data (Shih and Jones, 1989).

Table 8.1: Vapour-liquid equilibrium data for the propanoic acid (A) + pentanoic acid (B) system at 20 kPa.

T (K)	x_A	y_A	η_{A1}	η_{B1}	γ_A	γ_B
370.01	0.973*	0.992	0.4133	0.0058	0.9932	1.4286
370.82	0.955*	0.987	0.4186	0.0095	0.9864	1.3568
371.90	0.922	0.979	0.4251	0.0156	0.9856	1.2359
373.45	0.880	0.967	0.4342	0.0246	0.9804	1.1846
376.69	0.787	0.941	0.4523	0.0439	0.9828	1.0420
380.23	0.673	0.904	0.4664	0.0718	1.0093	0.9605
385.79	0.516	0.821	0.4709	0.1340	1.0399	0.9674
390.14	0.380	0.718	0.4477	0.2083	1.1131	0.9880
393.63	0.287	0.628	0.4182	0.2736	1.1863	0.9847
397.85	0.201	0.498	0.3604	0.3676	1.2270	1.0018
402.30	0.121	0.351	0.2772	0.4783	1.3101	0.9981
408.19	0.035*	0.133	0.1185	0.6530	1.5401	0.9932

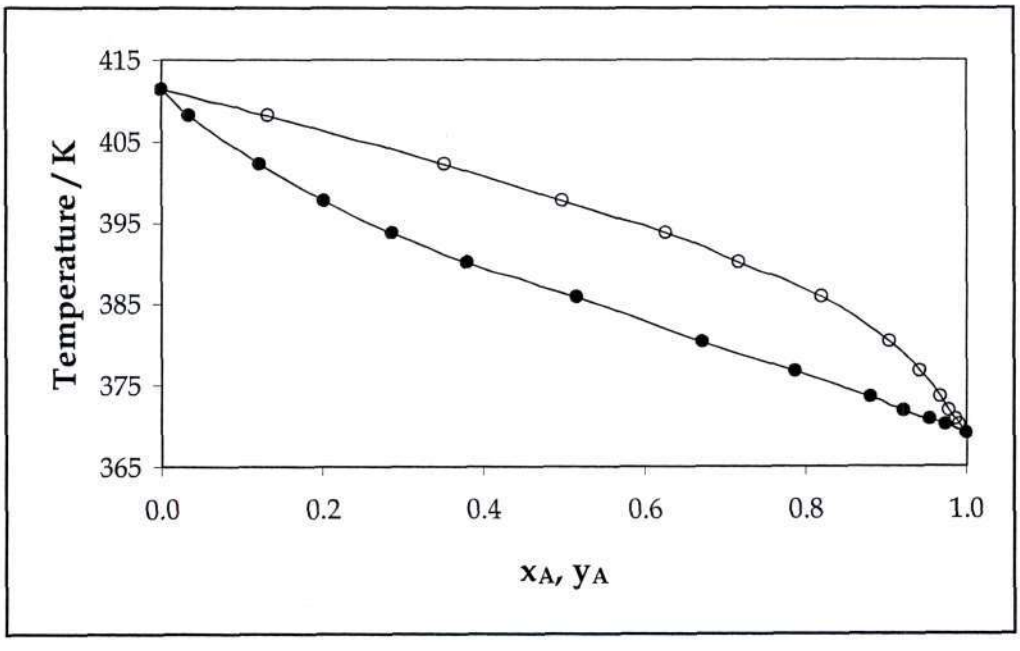

Figure 8.1: *T-x-y* curve for the propanoic acid (A) + pentanoic acid (B) system at 20 kPa. The circles are the experimental data and the solid line is a plot through these points.

Table 8.2: Vapour-liquid equilibrium data for the propanoic acid (A) + pentanoic acid (B) system at 393.15 K.

P (kPa)	x_A	y_A	η_{A1}	η_{B1}	γ_A	γ_B	$1/\Phi_A$	$1/\Phi_B$
10.78	0.025*	0.111	0.0975	0.6716	1.7838	0.9700	0.5660	1.0266
12.05	0.060*	0.235	0.1934	0.5711	1.6141	0.9566	0.5992	1.0393
15.21	0.151	0.457	0.3354	0.4013	1.4033	0.9399	0.6584	1.0538
18.73	0.264	0.615	0.4149	0.2844	1.2238	0.9459	0.7191	1.0595
24.49	0.418	0.761	0.4663	0.1767	1.1364	0.9715	0.7852	1.0574
31.56	0.588	0.859	0.4830	0.1038	1.0780	1.0391	0.8502	1.0494
37.51	0.726	0.916	0.4853	0.0621	1.0429	1.1109	0.9004	1.0405
43.13	0.849	0.956	0.4828	0.0328	1.0197	1.2248	0.9449	1.0313
46.92	0.926	0.978	0.4799	0.0164	1.0145	1.3039	0.9728	1.0249
48.16	0.965	0.990	0.4815	0.0074	0.9972	1.4085	0.9870	1.0215
48.76	0.980	0.994	0.4813	0.0046	0.9963	1.4266	0.9924	1.0202
49.48	0.990*	0.997	0.4804	0.0023	0.9984	1.4708	0.9960	1.0193

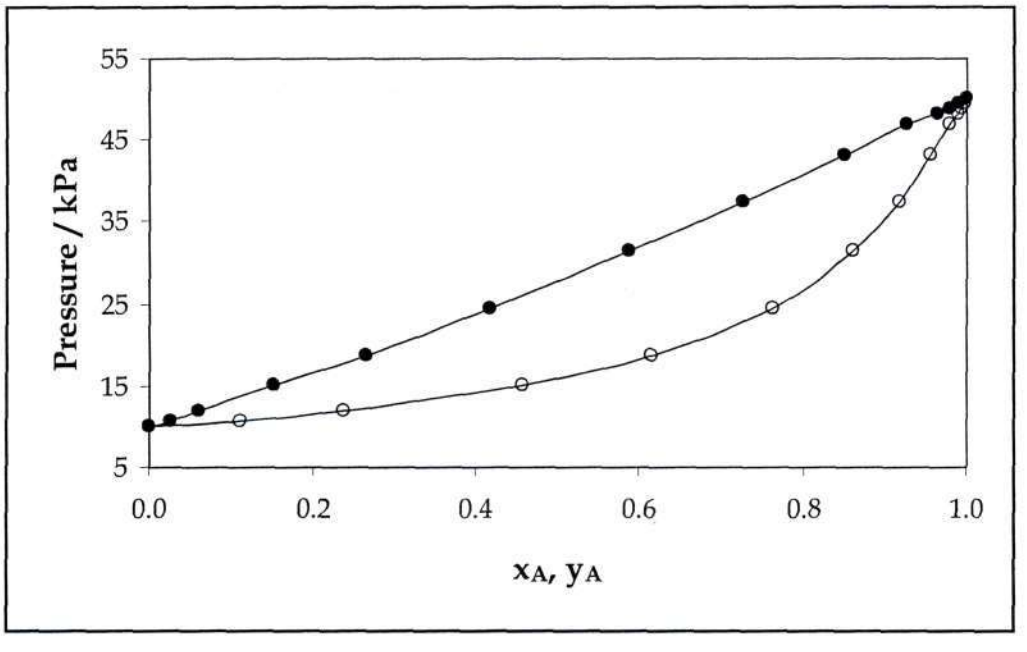

Figure 8.2: *P-x-y* curve for the propanoic acid (A) + pentanoic acid (B) system at 393.15 K. The circles are the experimental data and the solid line is a plot through these points.

Table 8.3: Vapour-liquid equilibrium data for the propanoic acid (A) + pentanoic acid (B) system at 403.15 K.

P (kPa)	x_A	y_A	η_{A1}	η_{B1}	γ_A	γ_B	$1/\Phi_A$	$1/\Phi_B$
15.23	0.005*	0.033	0.0307	0.7352	2.4455	1.0018	0.5540	1.0108
16.51	0.029	0.152	0.1330	0.6382	2.1002	0.9652	0.5823	1.0265
18.78	0.084	0.334	0.2699	0.4983	1.6668	0.9089	0.6341	1.0481
21.31	0.153	0.472	0.3590	0.3952	1.3829	0.8845	0.6765	1.0589
26.20	0.258	0.613	0.4327	0.2884	1.2161	0.9060	0.7262	1.0638
34.64	0.417	0.756	0.4866	0.1807	1.1199	0.9548	0.7906	1.0600
44.90	0.598	0.854	0.5050	0.1081	1.0494	1.0748	0.8582	1.0479
52.25	0.728	0.911	0.5123	0.0664	1.0179	1.1349	0.9045	1.0373
60.04	0.847	0.953	0.5124	0.0348	0.9963	1.2823	0.9464	1.0271
65.80	0.928	0.975	0.5086	0.0185	0.9986	1.5048	0.9746	1.0202
68.85	0.985*	0.994	0.5108	0.0043	0.9887	1.7417	0.9926	1.0157
69.87	0.994*	0.998	0.5100	0.0018	0.9927	1.8738	0.9964	1.0147

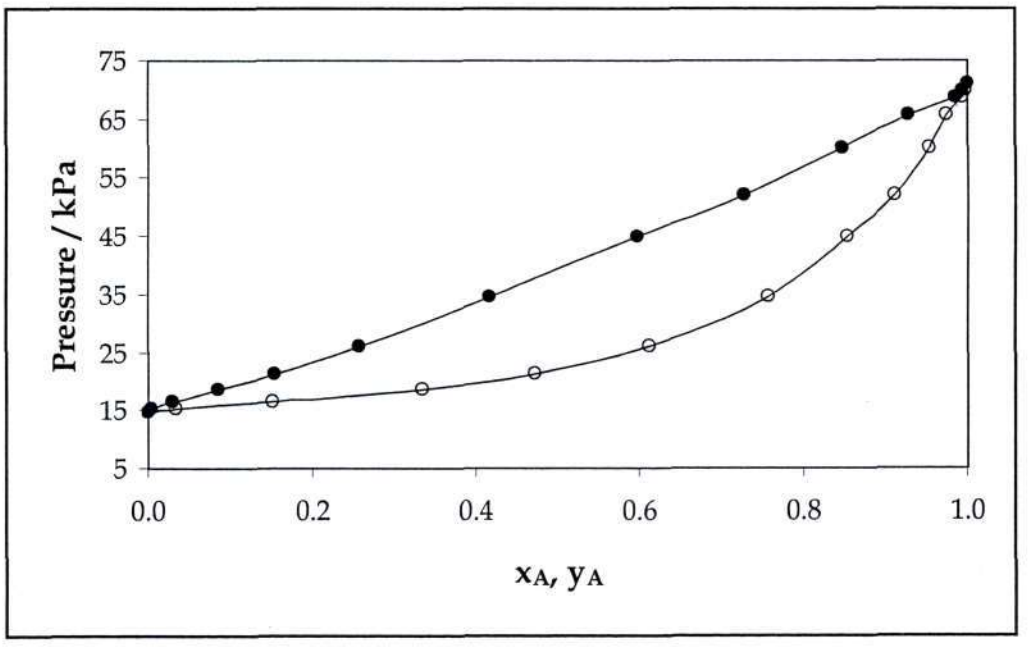

Figure 8.3: *P-x-y* curve for the propanoic acid (A) + pentanoic acid (B) system at 403.15 K. The circles are the experimental data and the solid line is a plot through these points.

Table 8.4: Vapour-liquid equilibrium data for the propanoic acid (A) + pentanoic acid (B) system at 413.15 K.

P (kPa)	x_A	y_A	η_{A1}	η_{B1}	γ_A	γ_B	$1/\Phi_A$	$1/\Phi_B$
22.24	0.007*	0.029	0.0277	0.7373	1.6590	1.0105	0.5709	1.0074
23.35	0.025*	0.093	0.0851	0.6837	1.4976	1.0019	0.5912	1.0185
27.33	0.086	0.255	0.2162	0.5506	1.2959	1.0075	0.6358	1.0371
30.85	0.147	0.382	0.3056	0.4536	1.2088	1.0039	0.6733	1.0469
38.41	0.261	0.568	0.4158	0.3159	1.1549	1.0042	0.7294	1.0524
48.71	0.403	0.714	0.4812	0.2093	1.0971	1.0440	0.7886	1.0487
54.24	0.483	0.769	0.5006	0.1689	1.0592	1.0842	0.8194	1.0441
65.90	0.634	0.861	0.5259	0.1023	1.0299	1.1275	0.8742	1.0336
78.93	0.794	0.930	0.5358	0.0523	1.0042	1.2245	0.9296	1.0221
88.90	0.897	0.967	0.5363	0.0251	1.0012	1.3304	0.9650	1.0149
96.10	0.986*	0.995	0.5386	0.0038	0.9897	1.5322	0.9929	1.0094
97.72	0.993*	0.998	0.5372	0.0018	0.9960	1.5788	0.9966	1.0087

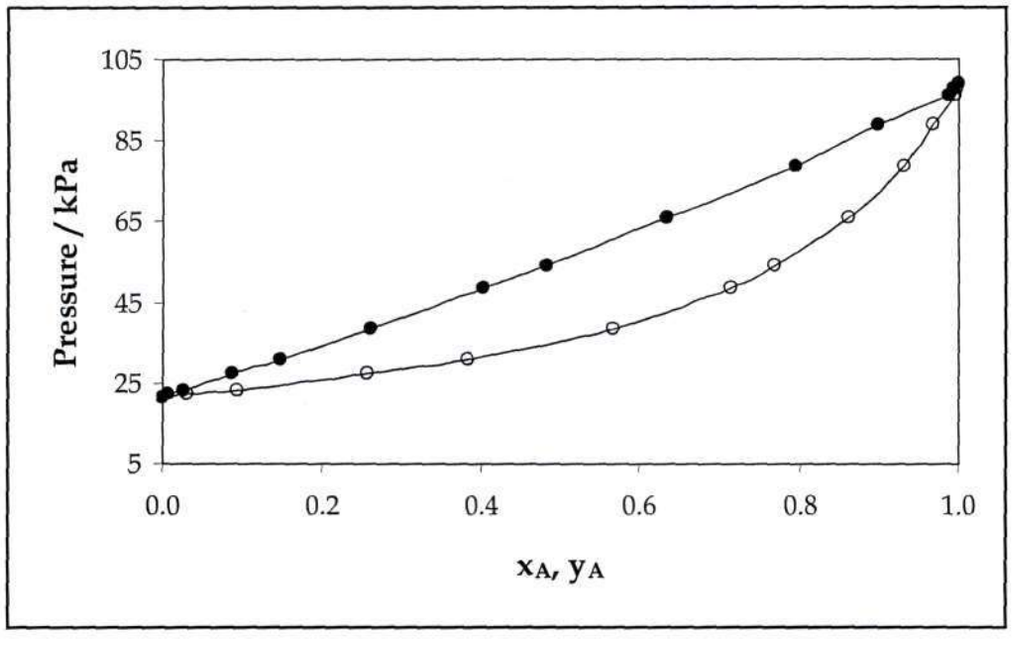

Figure 8.4: *P-x-y* curve for the propanoic acid (A) + pentanoic acid (B) system at 413.15 K. The circles are the experimental data and the solid line is a plot through these points.

Table 8.5: Vapour-liquid equilibrium data for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 20 kPa.

T (K)	x_A	y_A	η_{A1}	η_{B1}	γ_A	γ_B
382.66	0.990*	0.995	0.5987	0.0039	0.9941	1.4609
383.05	0.982	0.992	0.6001	0.0062	0.9867	1.3838
383.68	0.964	0.985	0.6007	0.0121	0.9791	1.3041
384.85	0.924	0.969	0.6001	0.0250	0.9713	1.1945
387.52	0.820	0.926	0.5937	0.0594	0.9670	1.0758
392.76	0.593	0.800	0.5502	0.1553	0.9984	1.0119
396.75	0.445	0.683	0.4972	0.2417	1.0235	0.9887
402.85	0.221	0.424	0.3423	0.4309	1.1180	0.9935
406.53	0.120	0.256	0.2218	0.5587	1.1589	0.9921
409.11	0.052*	0.123	0.1132	0.6642	1.2433	0.9933
410.48	0.022*	0.054	0.0515	0.7219	1.3002	0.9937
410.95	0.013*	0.034	0.0326	0.7395	1.3414	0.9917

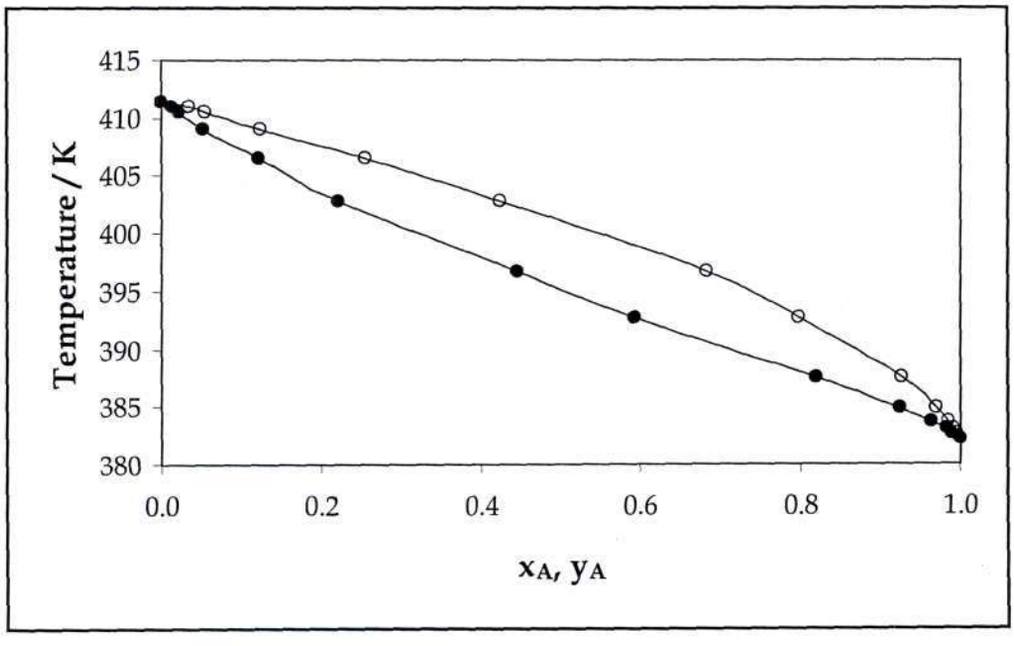

Figure 8.5: T-x-y curve for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 20 kPa. The circles are the experimental data and the solid line is a plot through these points.

Table 8.6: Vapour-liquid equilibrium data for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 393.15 K.

P (kPa)	x_A	y_A	η_{A1}	η_{B1}	γ_A	γ_B	$1/\Phi_A$	$1/\Phi_B$
10.12	0.009*	0.024	0.0230	0.7443	1.3247	0.9939	0.6400	1.0039
10.36	0.025	0.061	0.0574	0.7132	1.2685	0.9906	0.6504	1.0082
10.80	0.048	0.115	0.1046	0.6686	1.2382	0.9922	0.6670	1.0143
11.78	0.110	0.244	0.2104	0.5670	1.1957	0.9806	0.7063	1.0249
13.80	0.226	0.445	0.3542	0.4154	1.1448	0.9679	0.7665	1.0308
16.80	0.371	0.628	0.4631	0.2800	1.1092	0.9781	0.8250	1.0240
20.49	0.557	0.779	0.5365	0.1694	1.0449	1.0234	0.8846	1.0047
23.71	0.678	0.856	0.5640	0.1114	1.0438	1.0724	0.9185	0.9884
27.13	0.816	0.925	0.5849	0.0595	1.0287	1.1477	0.9544	0.9670
29.11	0.898	0.958	0.5938	0.0333	1.0189	1.2368	0.9748	0.9531
30.32	0.968	0.987	0.6043	0.0105	1.0020	1.2889	0.9921	0.9404
30.76	0.988*	0.995	0.6068	0.0039	0.9994	1.3328	0.9980	0.9358

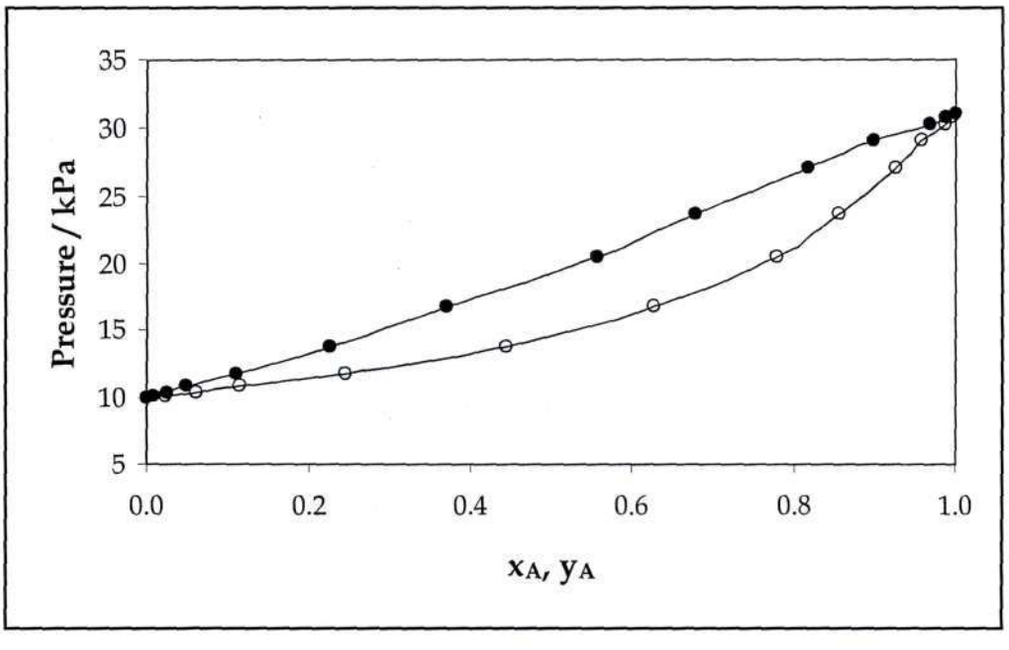

Figure 8.6: P-x-y curve for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 393.15 K. The circles are the experimental data and the solid line is a plot through these points.

Table 8.7: Vapour-liquid equilibrium data for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 403.15 K.

P (kPa)	x_A	y_A	η_{A1}	η_{B1}	γ_A	γ_B	$1/\Phi_A$	$1/\Phi_B$
14.94	0.009*	0.024	0.0227	0.7450	1.3312	0.9997	0.6454	1.0035
15.23	0.024	0.058	0.0545	0.7168	1.2612	0.9946	0.6554	1.0077
15.85	0.047	0.112	0.1026	0.6720	1.2196	0.9947	0.6717	1.0138
17.25	0.106	0.229	0.2003	0.5782	1.1643	0.9929	0.7080	1.0240
20.35	0.226	0.429	0.3454	0.4264	1.1115	0.9973	0.7670	1.0300
24.43	0.371	0.615	0.4603	0.2894	1.0837	0.9995	0.8229	1.0228
29.81	0.556	0.775	0.5414	0.1725	1.0373	1.0296	0.8819	1.0024
33.69	0.670	0.850	0.5718	0.1166	1.0265	1.0595	0.9144	0.9864
39.09	0.819	0.925	0.5955	0.0596	1.0153	1.1436	0.9539	0.9637
41.94	0.897	0.957	0.6037	0.0348	1.0086	1.2544	0.9739	0.9513
44.18	0.966	0.985	0.6128	0.0118	1.0006	1.3699	0.9915	0.9402
44.97	0.990*	0.996	0.6160	0.0036	0.9993	1.4449	0.9978	0.9362

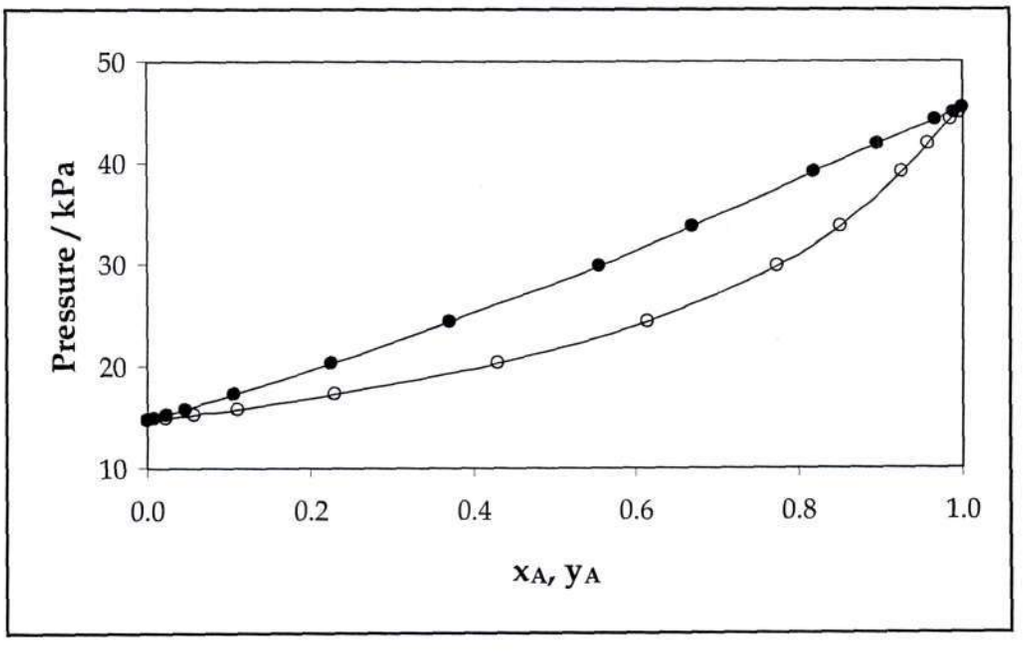

Figure 8.7: P-x-y curve for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 403.15 K. The circles are the experimental data and the solid line is a plot through these points.

Table 8.8: Vapour-liquid equilibrium data for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 413.15 K.

P (kPa)	x_A	y_A	η_{A1}	η_{B1}	γ_A	γ_B	$1/\Phi_A$	$1/\Phi_B$
21.70	0.008*	0.021	0.0199	0.7474	1.3123	1.0003	0.6521	1.0032
22.13	0.023*	0.054	0.0510	0.7198	1.2361	0.9971	0.6625	1.0075
23.07	0.048	0.107	0.0993	0.6747	1.1975	0.9996	0.6789	1.0136
25.00	0.104	0.216	0.1908	0.5875	1.1318	1.0030	0.7112	1.0226
29.43	0.222	0.405	0.3312	0.4421	1.0833	1.0233	0.7668	1.0283
34.79	0.367	0.590	0.4490	0.3070	1.0502	1.0322	0.8219	1.0209
42.47	0.554	0.764	0.5414	0.1809	1.0232	1.0544	0.8815	0.9993
48.90	0.673	0.847	0.5756	0.1184	1.0310	1.0849	0.9153	0.9820
54.37	0.772	0.903	0.5946	0.0764	1.0333	1.1132	0.9417	0.9671
59.33	0.873	0.944	0.6063	0.0447	1.0241	1.2158	0.9680	0.9516
61.88	0.927	0.969	0.6149	0.0251	1.0122	1.3060	0.9817	0.9435
64.24	0.991	0.996	0.6256	0.0032	1.0001	1.4291	0.9978	0.9340

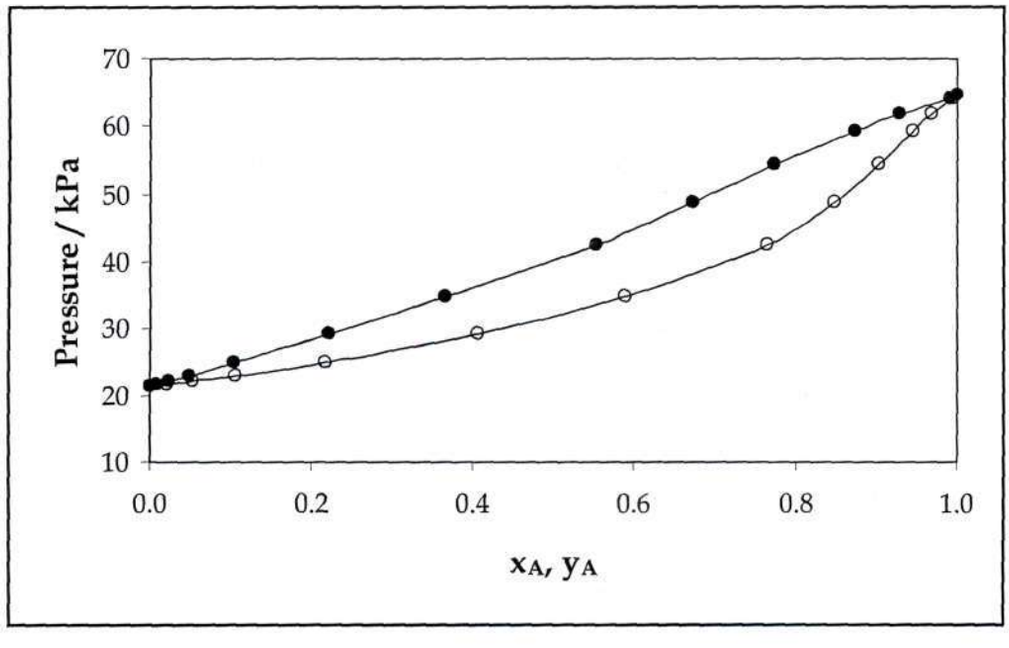

Figure 8.8: P-x-y curve for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 413.15 K. The circles are the experimental data and the solid line is a plot through these points.

The constants in the DIPPR equation (with K in kPa^{-1} and T in K):

$$\log_{10} K = A_K + \frac{B_K}{T} \tag{8.1}$$

are given for the three acids studied in Table A-1, Appendix A.

For both systems and for all temperatures and pressures, positive deviations from Raoult's law were found. Small negative deviations occurred, however, for the pentanoic acid activity coefficient, γ_B, in the pentanoic acid-rich regions, as shown in Figure 8.9 for the propanoic acid + pentanoic acid system at 403.15 K (the remaining liquid-phase activity coefficient plots are given in Appendix C). Tamir and Wisniak (1975) found positive deviations for the system formic acid + propanoic acid, but negative deviations for the acetic acid + propanoic acid system. A slight minimum was also found for γ_A in the propanoic acid-rich region of the isobaric systems. The mild minima observed for γ_B and γ_A (in the isobaric systems) were found to be characteristic of the carboxylic acid binary mixtures. This effect was seen to a greater or lesser degree

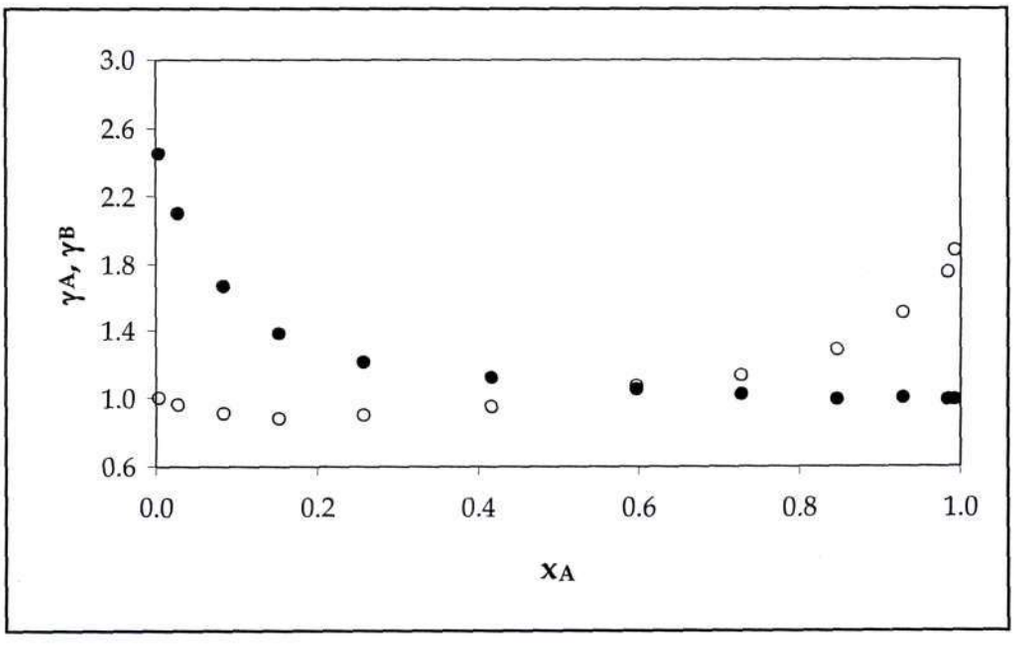

Figure 8.9: Liquid-phase activity coefficients, γ_A (filled symbols) and γ_B (open symbols) for the propanoic acid (A) + pentanoic acid (B) system at 403.15 K showing the small negative deviations that occurred in the pentanoic acid-rich region.

in the dilute regions of all the systems measured in the present study and are similar to those found by Tamir and Wisniak (1975) for their acetic acid + propanoic acid system at atmospheric pressure. Their activity coefficient data for the formic acid + propanoic acid system appear too scattered to perform a meaningful analysis or consistency test.

In Figure 8.10, an example is provided of typical true species vapour-phase concentration profiles as a function of the vapour-phase stoichiometric concentration, y_A, for the 2-methyl propanoic acid + pentanoic acid system at 393.15 K. In the chemical theory it is assumed that vapour-phase non-ideality arises only from the chemical species formation. As was discussed above, although both systems showed positive deviations from Raoult's Law, small negative deviations were observed in the dilute regions (particularly in the pentanoic acid-rich region). This effect is possibly due to appreciable heterodimer formation, illustrated in Figure 8.10.

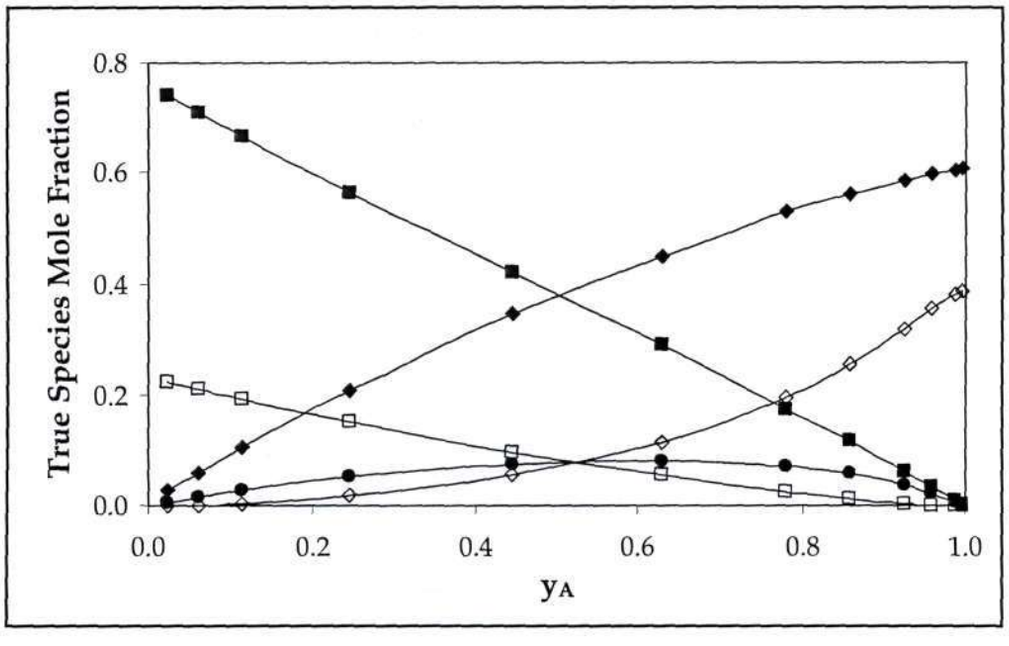

Figure 8.10: True species mole fractions versus stoichiometric vapour-phase mole fraction for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 393.15 K. Symbols are as follows: monomer A, filled diamonds; dimer A, open diamonds; monomer B, filled squares; dimer B, open squares and heterodimers AB, filled circles.

8.1.2 Thermodynamic Consistency Testing

The outcomes for the area test of thermodynamic consistency (Equation 2.22) for all the systems are shown in Table 8.9 and the associated plots are given in Appendix C, Section C.2. The results for the 2-methyl propanoic acid + pentanoic acid system are superior to those for the propanoic acid + pentanoic acid system, except at the highest temperature (413.15 K). The area residuals at 20 kPa and at 403.15 K for the 2-methyl propanoic acid system and 413.15 K for the propanoic acid system are remarkably small since they reflect not only experimental *P-T-x-y* errors, but also those residing in all three K values. The reason for the higher residuals (in general) calculated for propanoic acid is not clear, but could be due to the formation of small amounts of trimer or tetramer, as suggested by Sebastiani and Lacquaniti (1967) for systems involving acetic acid.

It should also be noted that the result of the area test depends strongly on the total value of the area under the $\ln(\gamma_A/\gamma_B)$ vs. x_A curve. In this case, the systems do not deviate strongly from ideal, Raoult's law-type behaviour. Hence, a rather high area deviation (integral value divided by total area) can be produced by much smaller experimental errors than for a more non-ideal system. The inability of the area test to take into account measured pressures was not considered serious since previous vapour pressure measurements (Clifford et al., 2004) using the same equipment and procedures gave very satisfactory results.

Table 8.9: Results from the area test of thermodynamic consistency.

Temperature/ Pressure	Propanoic Acid + Pentanoic Acid	2-methyl propanoic Acid + Pentanoic Acid
	ΔA (%)	ΔA (%)
20 kPa	44.60	7.26
393.15 K	47.27	27.08
403.15 K	33.67	3.93
413.15 K	9.47	27.44

8.1.3 Distillation Equipment Design for Carboxylic Acid Mixtures

Distillation equipment design requires that the activity coefficients γ_A and γ_B be represented as functions of composition by well-known equations for the excess Gibbs energy (such as the Wilson (Wilson, 1964), NRTL (Renon and Prausnitz, 1968) or UNIQUAC (Abrams and Prausnitz, 1975) equations mentioned in the introduction to Section 8.1). To allow effective process design at intermediate temperatures, i.e. at temperatures other than those covered in the isothermal experiments, the γ_i must also be known as functions of temperature (usually through the use of a suitable empirical correlation). In addition, a bubble-point (commonly termed a bubble P) calculation must be performed to determine the P-x_A profile at the chosen temperature. This procedure for systems involving dimerizing components differs somewhat from that normally employed. The technique employed in this study (discussed in Section 2.4 of Chapter 2) also yields the required vapour composition at any arbitrary liquid composition, x_A.

Experimentation with well-known equations for the excess free energy (like those given above) showed the T-K Wilson equation (Tsuboka and Katayama, 1975) to be the most suitable for the carboxylic acid systems measured, as judged by the area test (Equation 2.22). The regression procedure that was used to determine the T-K Wilson parameters for the carboxylic acid systems studied here was outlined in Section 2.4. These parameters are presented in Table 8.10 and are plotted as functions of temperature in Figures 8.11 and 8.12. They displayed a quadratic dependence upon temperature:

$$\lambda_{ij} = aT^2 + bT + c \tag{8.2}$$

and the constants associated with Equation 8.2 for both systems are given in Table 8.11.

The average pressure and vapour-phase residuals (δP and δy_A) calculated using this procedure are presented for the isothermal runs in Table 8.10. The predictions are very good for the 2-methyl propanoic acid + pentanoic acid system (as is illustrated by the example given in Figure 8.13) and satisfactory for the propanoic acid + pentanoic acid system. Section C.3 of Appendix C contains the P-x-y and x-y plots for the remaining isothermal systems.

The values of the vapour-phase correction factors, Φ_i (Equation 2.28), extend across a wide range (from less than 0.60 to more than 1.00; see Tables 8.1 to 8.8). This confirms the considerable departure from ideal gas behaviour that occurs in the vapour phase of carboxylic acid systems, and the necessity of including theory that adequately incorporates effects like dimerization.

Table 8.10: T-K Wilson equation parameters fitted to the experimental liquid-phase activity coefficients.

T (K)	Propanoic Acid + Pentanoic Acid			
	λ_{12} (J/mol)	λ_{21} (J/mol)	δP (kPa)	δy_A
393.15	5300.12	-2970.52	0.645	0.017
403.15	7468.77	-3687.04	0.877	0.019
413.15	6333.94	-3859.13	0.544	0.008
T (K)	2-methyl propanoic Acid + Pentanoic Acid			
	λ_{12} (J/mol)	λ_{21} (J/mol)	δP (kPa)	δy_A
393.15	1367.73	-567.01	0.304	0.005
403.15	2948.18	-1821.28	0.225	0.004
413.15	3932.47	-2445.54	0.620	0.003

Table 8.11: Quadratic equation (Equation 8.2) coefficients for the temperature dependence of the T-K Wilson equation parameters (range: 393.15 K - 413.15 K).

T-K Wilson Parameter	Propanoic Acid + Pentanoic Acid			2-methyl propanoic Acid + Pentanoic Acid		
	a	b	c	a	b	c
λ_{12}	-16.5174	13369.67	-2697942	-2.9808	2531.64	-533217
λ_{21}	2.7222	-2239.34	456663.8	3.1500	-2633.81	548022.3

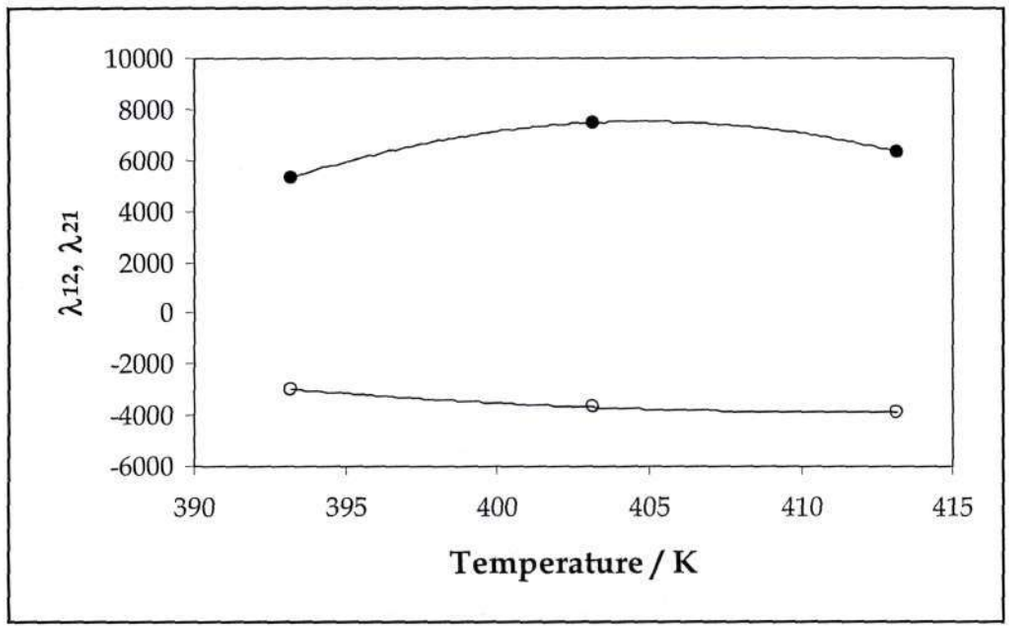

Figure 8.11: Temperature dependence of the T-K Wilson parameters, λ_{ij}, for the propanoic acid (A) + pentanoic acid (B) system. Filled symbols represent λ_{12}, open symbols λ_{21}.

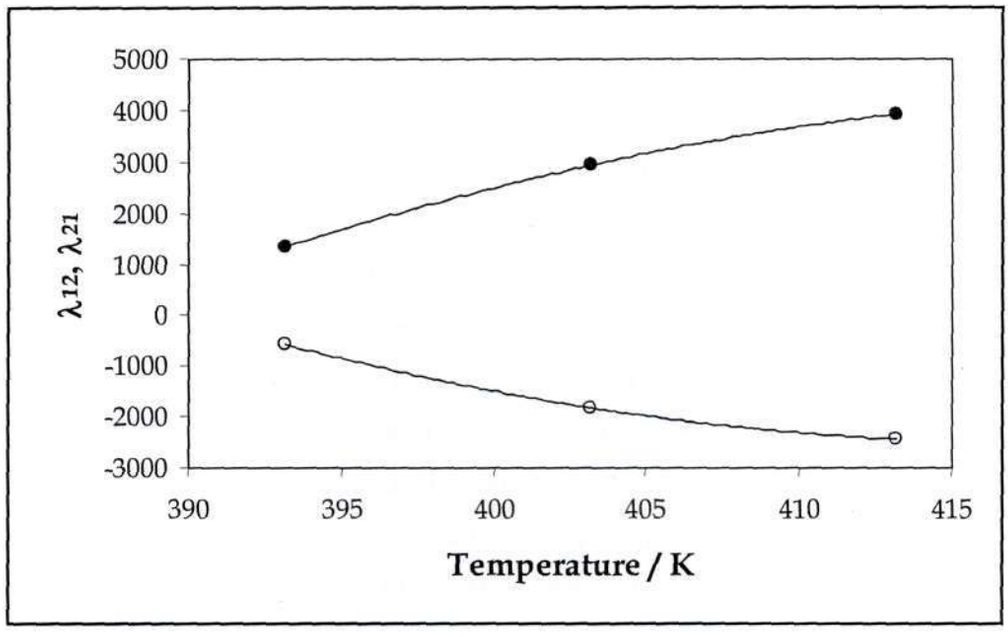

Figure 8.12: Temperature dependence of the T-K Wilson parameters, λ_{ij}, for the 2-methyl propanoic acid (A) + pentanoic acid (B) system. Filled symbols represent λ_{12}, open symbols λ_{21}.

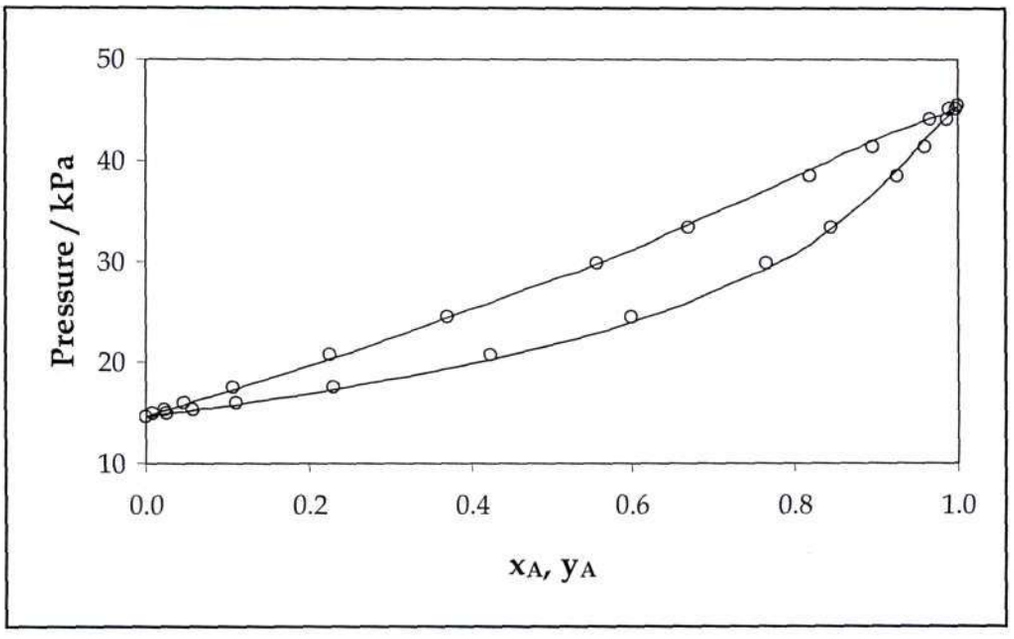

Figure 8.13: Example of the predictions that may be achieved using the proposed bubble point calculation scheme (2-methyl propanoic acid (A) + pentanoic acid (B) at 403.15 K). The solid lines are the experimental data and the open symbols are the model prediction.

8.2 Molecular Simulation Results

The molecular simulation results are divided into several sections. The first section deals with the simulation details and adjustable settings that were common to all the simulations run during this project (for example, the Ewald summation parameters). Thereafter, the actual simulation results will be presented.

The pure component simulation results are given first and consist of simulations involving alcohols and carboxylic acids. The data presented in this section have, for the most part, been simulated previously in the literature, and were used primarily to gain familiarity with the techniques involved in simulating polar compounds. Thus, they paved the way for the more complex and involved simulations that were to follow.

The next set of results that will be presented are those from the parameterization of H_2S, SO_2 and NO_2. This includes both the pure component simulation data generated during the fitting process and the principal focus of this section: the parameters themselves.

This is followed by the binary phase equilibrium (VLE) results. The carboxylic acid systems studied were propanoic acid + pentanoic acid and 2-methyl propanoic acid + pentanoic acid (an isotherm and an isobar were simulated for each system). Thereafter, the parameters developed in the preceding section for hydrogen sulfide (H_2S) and sulfur dioxide (SO_2) were used (in conjunction with the TraPPE-UA force field) to simulate binary VLE data with certain short-chain alkanes and 1-alcohols.

A section is also devoted to an investigation into the sensitivity of the Lennard Jones parameters of the force fields used in this work and, in particular, the combining rules used to determine unlike interaction parameters. A sensitivity analysis was carried out during the parameterization of H_2S, SO_2 and NO_2, and the effect of different combining rules on the binary VLE simulation predictions was also examined.

The final results that will be presented pertain to association-type effects such as hydrogen bonding and dimerization. Molecular simulation provides an ideal tool for viewing the microscopic structure of chemical systems and allows for determination of, for example, the extent to which dimerization takes place in the vapour phase of the carboxylic acid VLE systems through the analysis of radial distribution functions. Hence, the accuracy of the largely empirical equation to determine the heterodimer chemical equilibrium constant, K_{AB}, can be ascertained.

8.2.1 Simulation Details

The simulation details for all of the simulations run during this project are discussed here. Those that were common to every simulation are presented first, followed by sections dealing with those that varied depending upon the system and type of simulation being conducted (e.g. number of molecules).

8.2.1.1 Common Simulation Settings

All of the simulations conducted during this project were Monte Carlo simulations in the Gibbs Ensemble (discussed in detail in Chapter 6), as implemented in the MCCCS Towhee source code (Martin, 2005). The sampling of phase space was enhanced using the coupled-decoupled configurational bias method (Martin and Siepmann, 1999). An

inner cutoff radius of 10 Å was used for the intermolecular interactions following the dual-cutoff algorithm of Vlugt et al. (1998), with the bias removed through application of analytical tail corrections in the acceptance criteria. Use of the dual-cutoff technique more than doubles the configurational bias Monte Carlo (CBMC) computation speed (Martin, 2005). For the Lennard-Jones part of the potential, a spherical potential truncation was enforced for pseudo-atom pairs separated by an intermolecular distance exceeding 14 Å. Again, any bias that was introduced was removed through analytical tail corrections. The Ewald summation method (Allen and Tildesley, 1987; Frenkel and Smit, 2002) with tinfoil boundary conditions (i.e. the surroundings are assumed to have an infinite dielectric constant (Aguado and Madden, 2003)) was used to calculate the long-range electrostatic energy. The maximum number of inverse space vectors, K_{max}, was set to 5 and alpha was calculated by dividing κL (set to 5.6) by the shortest box length (Allen and Tildesley, 1987; Martin, 2005).

Further details regarding the configurational bias Monte Carlo technique and the Ewald summation may be found in Sections 4.2 and 5.4, respectively.

8.2.1.2 Number of Monte Carlo Steps

Any Gibbs ensemble simulation consists of two parts: the first is termed the equilibration stage and the second the production stage. The equilibrium stage is used (as the name suggests) to bring the system to a state of equilibrium. This is determined by periodically checking the simulation output until the running averages of important system properties (for example, the saturated densities) have leveled out and the property values are only varying within a statistically acceptable range i.e. the standard deviation is less than a predetermined value. Generally, this is considered to be approximately 1 % for the saturated liquid density and 5-10 % for the saturated vapour density and pressure. However, at higher temperatures the standard deviations may be larger since the increased system energy results in unlikely system configurations being accepted. During the equilibrium portion of the Markov chain (simulation), certain simulation settings such as the maximum translational, rotational and volume displacements are automatically updated in an attempt to achieve the target acceptance rates (discussed further in Section 8.2.1.4 below). In Figure 8.14, an example is given of a simulation (1-pentanol at 450 K) attaining equilibrium. It is

evident that this has occurred after approximately 25 000 Monte Carlo cycles, and (in practice) the production stage would run over the period from 25 000 to 50 000 cycles.

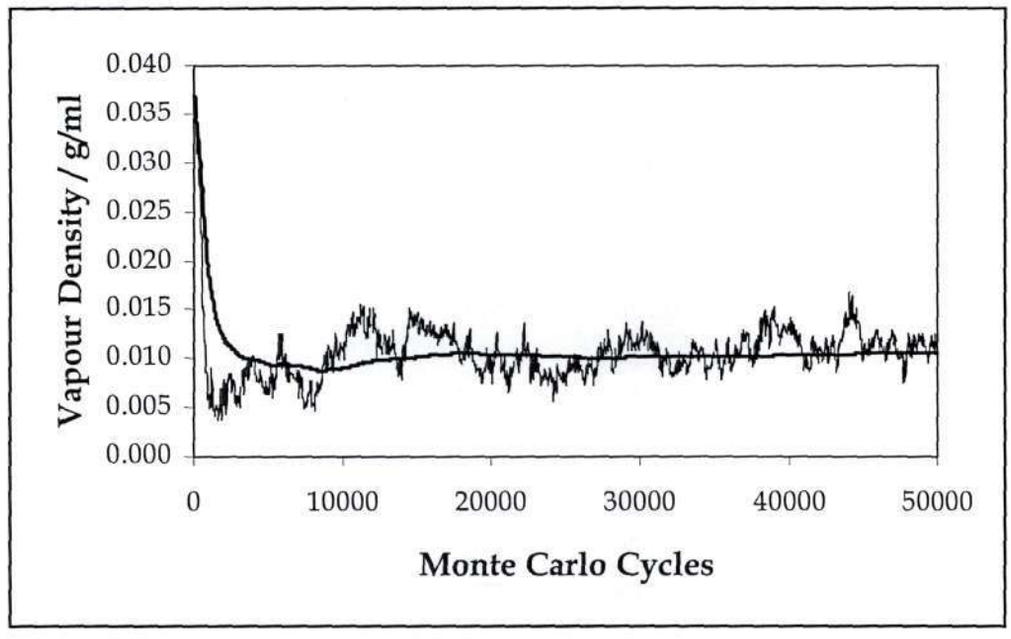

Figure 8.14: Plot of the vapour density of 1-pentanol at 450 K showing the system reaching equilibrium after approximately 25 000 Monte Carlo cycles. The dashed line represents the individual vapour density values and the solid line is the running average.

Once the system has reached equilibrium, the production stage of the simulation is initiated. During these production periods, run statistics are recorded that allow the ensemble averages such as pressure and density to be calculated. In order to determine the standard deviations associated with each ensemble property, the production cycle is divided into five equally-sized blocks. By studying the behaviour of the block averages, the statistical errors may be computed (Frenkel and Smit, 2002). It is important to ensure that the production stage is of a sufficient length to generate blocks large enough to yield a reliable estimate for a particular quantity. On the other hand, simulations of polar molecules are computationally demanding and often take days to complete (refer to Table 8.12 below). A happy medium must be struck between these two concerns that permits statistically converged data to be calculated in a computationally feasible time.

Table 8.12: The computational time required for the various simulations run during this project.

System	Time/Cycle (min)	Time/Point (h)	Points	Time/System (days)	System Size
Methanol	0.073	72.7	3	9.09	500
Propanol	0.084	69.6	5	14.50	250
2-methyl-1-propanol	0.082	75.3	3	9.41	200
2-methyl-2-propanol	0.088	80.5	5	16.77	200
Pentanol	0.108	89.7	5	18.69	150
Acetic Acid	0.226	84.6	5	17.63	300
Propanoic Acid	0.312	78.1	5	16.27	250
2-methyl propanoic Acid	0.215	53.7	5	11.19	200
Pentanoic Acid	0.247	61.8	5	12.88	200
H_2S	0.119	29.8	6	7.45	450
SO_2	0.079	19.8	5	4.13	400
NO_2	0.108	27.1	5	5.65	500
Propanoic Acid + Pentanoic Acid	0.246	82.1	14	47.89	300
2-methyl propanoic Acid + Pentanoic Acid	0.278	92.8	14	54.13	300
H_2S + Methane	0.143	71.6	7	20.88	800
H_2S + Ethane	0.148	61.5	6	15.38	800
H_2S + Propane	0.151	50.4	6	12.60	800
H_2S + Methanol	0.129	32.3	6	8.08	800
SO_2 + Methane	0.136	45.4	6	11.35	800
SO_2 + Propane	0.158	39.5	6	9.875	800
SO_2 + Ethanol	0.365	91.2	6	22.8	800

It is important to note that the information contained in Table 8.12 represents the best-case scenario i.e. the times shown are for the fastest simulations (those at the highest temperature) in each system. Also, the time consumed by simulations that failed or crashed is not included. For the parameterization work, more than 300 sets of parameters were tested in total. This amounts to (approximately) 530 days of CPU time. It is clear that work of this nature is only feasible if the researcher has access to a

Beowulf cluster or some other resource for performing numerous computations simultaneously.

Another point that should be made is that the Beowulf cluster was upgraded a year into the project and the new nodes resulted in improved simulation times (± 50 %). This is evident if one compares the time required for the 2-methyl propanoic acid pure component simulations using the new nodes (53.7 hours per point) to that of the old nodes (80.0 hours). Furthermore, the time necessary for the propanoic acid simulations was 78.1 hours, which is almost half again as long as those for 2-methyl propanoic acid. Normally, one would anticipate that simulations involving a larger molecule would take longer, since the computational complexity of the system has increased. The times given in Table 8.12 are, wherever possible, for simulations performed on the upgraded nodes. The Beowulf cluster upgrade and the installation of the new nodes are discussed further in Section 7.3.3.

For the alcohol pure component simulations the number of Monte Carlo steps were, in general, the same for both the equilibration and production cycles and varied between 3.75 and 15 million depending upon the particular alcohol being simulated. The carboxylic acid stages (for the pure component and binary VLE simulations) differed to some extent, with equilibration of the systems (as evidenced by constant average vapour and liquid densities) obtained by running the simulations for between 1.5 and 5 million Monte Carlo steps. The production periods during which ensemble data were recorded consisted of 3 to 6 million Monte Carlo steps. For the parameterization of H_2S, SO_2 and NO_2, equilibration was attained after running the simulations for 2 to 2.5 million Monte Carlo steps, followed by a production stage consisting of 4 to 5 million steps. These simulation lengths are consistent with those used by Nath when parameterizing H_2S (Nath, 2003). The binary VLE simulations involving H2S and SO2 required equilibrium and production stages of 4 to 8 and 8 to 16 million Monte Carlo steps, respectively.

The time taken for an individual simulation to reach equilibrium varied from system to system and, on occasion, even within a system. An effort was made to maintain a constant simulation length for the production stages of each system to ensure the maximum possible consistency in the calculated statistical errors.

8.2.1.3 *System Sizes*

The system size refers to the number of molecules present in that system. In any simulation, this is a significant decision and requires some thought. The computational intensity of a simulation (particularly one involving polar compounds in which charged interactions become critically important) scales geometrically with increased system size. Thus, the required simulation time increases appreciably as the number of molecules increases. At the same time, the simulation results are based on ensemble averages and, hence, the more molecules within the system, the smaller the thermal fluctuations will be (permitting the accumulation of accurate ensemble averages from less cycles i.e. shorter runs). This is especially pertinent for the vapour phase, in which there are often fewer molecules to contribute toward the system averages than is the case for the liquid phase. Furthermore, if too few molecules are present in a simulation, finite-size effects (which should be avoided) begin to influence the results.

Another reason for careful selection of the system size is that the box volumes are critically dependent upon the number of molecules present in the system. Typically, this is not a concern for the vapour boxes, which tend to be large in comparison to the liquid boxes. However, if the initial liquid box volume was set too low, the liquid box lengths frequently fell below the minimum image limit of twice the potential truncation (14 Å). If this occurs, the simulation has essentially failed and the result should generally be discarded and the simulation re-initiated with the liquid box volume increased. The most effective way to achieve this is to increase the system size. In general, the number of molecules for a particular system was set so as to yield liquid box lengths between 30 and 35 Å, which led to converged ensemble averages.

The system sizes used during this project are given in Table 8.12 above.

8.2.1.4 *Fixed Probabilities*

The fixed probabilities for the various types of moves in each simulation differed slightly depending on the particular system studied (but were kept constant for a particular system or class of compound e.g. alcohols). The probability of performing CBMC particle moves was adjusted to obtain approximately one accepted molecule swap and one volume exchange per ten Monte Carlo cycles. The fixed probabilities

given below and the maximum volume, translation and rotational displacements were selected so that an acceptance ratio of approximately 50 % was attained for each of these moves. These target values were chosen largely because they are the accepted figures quoted in the majority of the literature. However, they are, in reality, based on little more than empirical experience. Some researchers, for instance, use an acceptance ratio of 33 % for the volume, translational and rotational moves. In the end it amounts to personal preference. The fundamental consideration when setting the fixed probabilities and acceptance criteria is whether sufficient molecule swap and volume exchange moves are being performed to achieve chemical and mechanical equilibrium.

For all of the pure component simulations the probabilities were as follows: volume change, 0.006; CBMC, 0.194; molecule regrowth, 0.267; molecule translation, 0.266; molecule rotation, 0.267. The aggregated volume bias (AVB) move was included for the carboxylic acid binary VLE simulations to help prevent the strongly associating and cross-associating acids from forming clusters (this was not a major concern for the pure component acid simulations). Thus, the fixed probabilities for these binary simulations were: volume change, 0.002; CBMC, 0.198; AVBMC, 0.050; molecule regrowth, 0.250; molecule translation, 0.250; molecule rotation, 0.250. For the remaining binary systems (those involving H_2S and SO_2), the AVBMC move was removed since molecular clusters did not present a problem for these compounds: volume change, 0.002; CBMC, 0.198; molecule regrowth, 0.267; molecule translation, 0.266; molecule rotation, 0.267.

8.2.2 Pure Component Simulation Results

This section presents the results of the pure component simulations carried out for the following alcohols and carboxylic acids: methanol, 1-propanol, 2-methyl-1-propanol, 2-methyl-2-propanol, 1-pentanol, acetic acid, propanoic acid, 2-methyl propanoic acid and pentanoic acid. The data for 2-methyl-1-propanol, propanoic acid and 2-methyl propanoic acid represent new simulation predictions, whilst the other alcohols and acids have been simulated previously by Chen et al. (2001) and Kamath et al. (2004), respectively.

These initial simulations served a dual purpose: they provided an opportunity to become familiar with the intricacies of running successful simulations using the MCCCS Towhee source code and also ensured that the methods and software being

employed were capable of replicating the results achieved by other researchers. The latter is of particular importance as it establishes confidence in the validity and accuracy of the results that followed.

8.2.2.1 Alcohols

The TraPPE-UA force field was used to predict saturated vapour pressure and density data for the five alcohols mentioned above, which included both linear and branched compounds. The simulation results are given in Table 8.13 and are plotted in Figures 8.15 to 8.18, along with the applicable simulation data of Chen et al. (2001). All of the experimental data used for comparison were taken from Smith and Srivastava (1986).

The simulation results for the saturated vapour pressures displayed an average standard deviation of 8.1 % and are plotted in Figures 8.15 and 8.16. Clausius-Clapeyron plots were used to allow the normal boiling point and critical pressure predicted by the simulation data to be determined (discussed further below). Chen et al. (2001) state that the TraPPE-UA alcohol and alkane force fields over-predict pure component vapour pressures. The simulation results presented here confirm this and show that the TraPPE-UA force field over-predicts alcohol vapour pressures by less than 20 % on average, with the deviation from experimental data decreasing as the temperature increases. It should be noted, however, that this effect only becomes pronounced as the molecules increase in size. For low molecular weight alcohols such as methanol the vapour pressure prediction is much better (this is illustrated in Figure 8.15), with an average deviation from measured data of just over 5 %. Also, for methanol the vapour pressure is under-predicted at low temperatures.

It is evident from Figures 8.17 and 8.18 that the saturated densities agree well with experiment and the previously published simulation data of Chen et al. (2001). The saturated liquid densities differ from experiment by approximately 2 % (on average) and from the data of Chen and co-workers by less than 0.75 %, with a mean standard deviation of 1.4 %. The saturated vapour densities were, as expected, over-predicted by the TraPPE-UA force field and had an average standard deviation of 9.5 %. The vapour-liquid coexistence curves shown in Figures 8.17 and 8.18 were used to calculate the critical temperature and density from the simulation results by fitting to the density

scaling law and law of rectilinear diameters (Rowlinson and Swinton, 1982). Chen et al. (2001) found that for alcohols, the different shape of the coexistence curves (as opposed to non-polar alkanes) necessitated values of the scaling exponent that were smaller than the universal Ising value of 0.325. Therefore, values of 0.28 for methanol and 0.29 for heavier alcohols (the effective scaling exponents (β_{eff}) used by Chen et al. (2001) in their work) were used during the fitting procedure. The critical temperature and density calculated in this manner are presented in Table 8.14, along with the critical pressures and normal boiling points obtained from respective extrapolation and interpolation of the Clausius-Clapeyron plots.

Table 8.13: Numerical results for the pure component alcohol simulations.

Compound	T (K)	P (kPa)	ρ_{liq} (g/ml)	ρ_{vap} (g/ml)
Methanol	300	16.80	0.783	2.25E-04
	375	368.05	0.692	4.29E-03
	475	4359.14	0.508	6.77E-02
1-propanol	300	3.17	0.791	7.47E-05
	350	46.60	0.744	1.01E-03
	400	344.35	0.693	7.18E-03
	450	1254.55	0.623	2.59E-02
	500	3309.67	0.518	7.81E-02
2-methyl-1-propanol	400	254.86	0.691	6.07E-03
	450	913.54	0.621	2.16E-02
	500	2350.34	0.523	5.99E-02
2-methyl-2-propanol	300	4.69	0.750	7.47E-05
	350	121.97	0.728	3.27E-03
	400	637.68	0.655	1.65E-02
	450	1710.57	0.554	4.62E-02
	480	3234.84	0.465	1.05E-01
1-pentanol	300	0.42	0.810	1.50E-05
	350	9.44	0.770	2.90E-04
	400	81.26	0.721	2.31E-03
	450	409.25	0.661	1.10E-02
	500	1250.42	0.591	3.46E-02

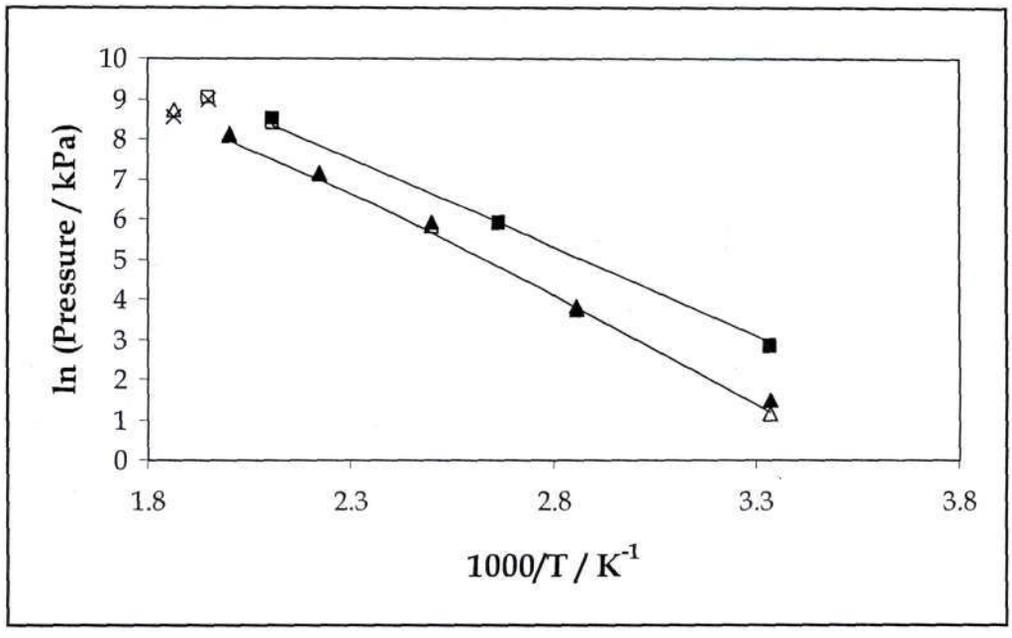

Figure 8.15: Clausius-Clapeyron plot for methanol (□) and 1-propanol (Δ). Experimental data are represented by solid lines and crosses (critical point) and symbols represent simulation results. Filled symbols are Chen et al. (2001) data.

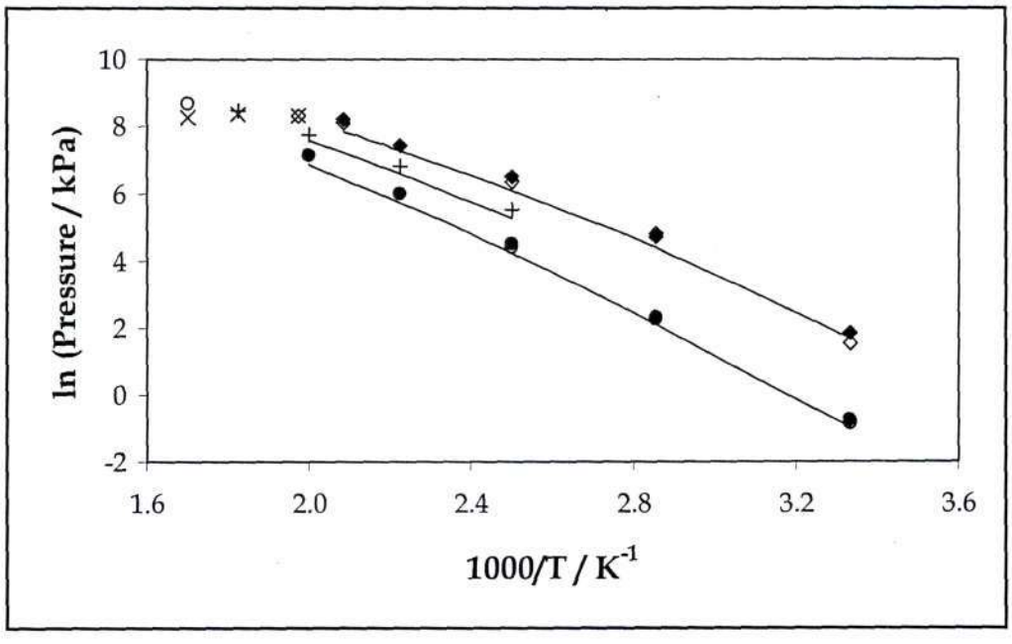

Figure 8.16: Clausius-Clapeyron plot for 2-methyl-1-propanol (+), 2-methyl-2-propanol (◊) and 1-pentanol (○). Experimental data are represented by solid lines and crosses (critical point) and symbols represent simulation results. Filled symbols are Chen et al. (2001) data.

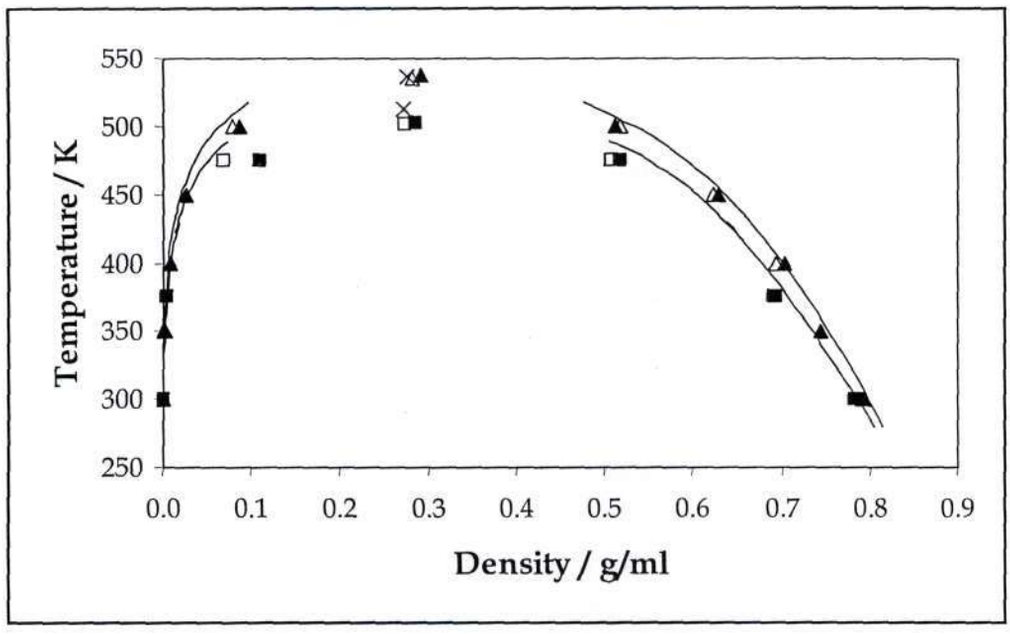

Figure 8.17: Vapour-liquid coexistence plot for methanol (□) and 1-propanol (Δ). Experimental data are represented by solid lines and crosses (critical point) and symbols represent simulation results. Filled symbols are Chen et al. (2001) data.

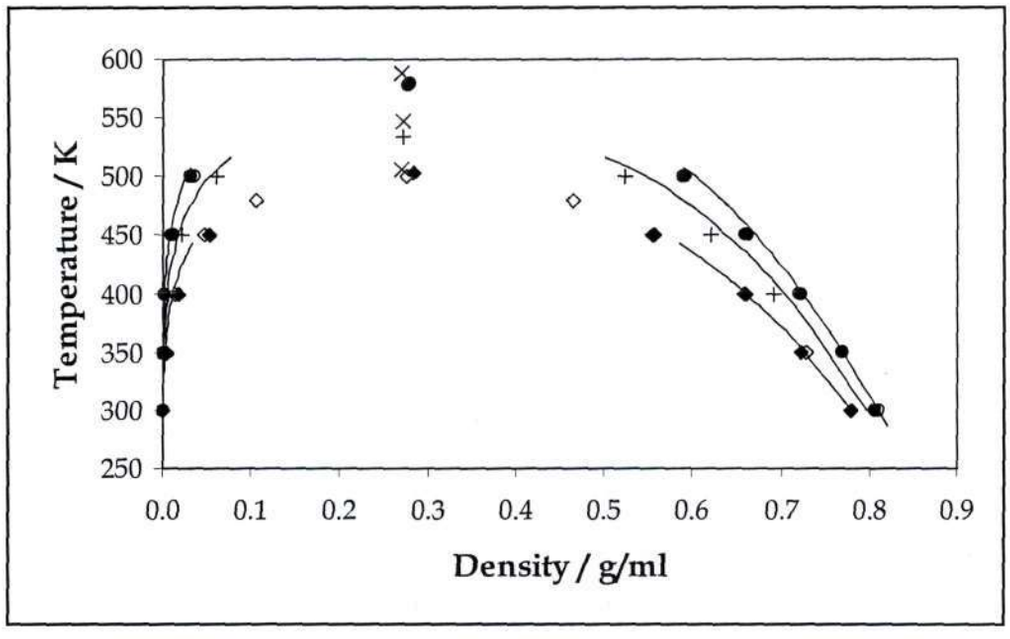

Figure 8.18: Vapour-liquid coexistence plot for 2-methyl-1-propanol (+), 2-methyl-2-propanol (◊) and 1-pentanol (○). Experimental data are represented by solid lines and crosses (critical point) and symbols represent simulation results. Filled symbols are Chen et al. (2001) data.

Table 8.14: Critical properties and normal boiling points for the alcohols.

Compound	Data Source	T_c (K)	ρ_c (g/ml)	P_c (kPa)	T_b (K)
Methanol	Experiment	512.64	0.272	8092	337.70
	Chen et al.	502	0.285		340
	This work	500.75	0.273	8436	338.37
1-propanol	Experiment	536.78	0.275	5170	370.29
	Chen et al.	538	0.290		368
	This work	535.31	0.282	6053	370.88
2-methyl-1-propanol	Experiment	547.78	0.272	4300	381.03
	This work	534.81	0.271	4706	368.99
2-methyl-2-propanol	Experiment	506.21	0.270	3973	355.49
	Chen et al.	504	0.283		354
	This work	500.83	0.274	4147	357.08
1-pentanol	Experiment	588.15	0.270	3909	411.16
	Chen et al.	579	0.279		406
	This work	577.76	0.277	5666	409.13

The critical temperatures are slightly under-estimated in all cases when compared to experiment, with the average discrepancy under 1.5 % and the estimated critical densities are generally over-predicted, with the average deviation being 1.7 %. This compares favourably with the 3 % deviation quoted by Chen and co-workers. The over-estimation of the critical densities is attributed primarily to the fact that the law of rectilinear diameters has a negative slope and, hence, the under-prediction of the critical temperatures results in an over-prediction of the associated densities (Chen et al., 2001). Also, the TraPPE-UA force field yields values for the saturated vapour densities that are too large, particularly at higher temperatures. Naturally, this leads directly to inflated critical densities.

The interpolated normal boiling points (T_b) are in good agreement with measured data, differing by less than 0.5 %, except for 2-methyl-1-propanol where the boiling point is under-estimated by approximately 3 %. For all the alcohols, the estimated critical pressures are too large and deviated from experiment by an average of 16.0 %. This value is skewed to some extent by the fact that the critical pressure of 1-pentanol

is over-predicted by almost 45 %. In other words, the critical pressure of the remaining four alcohols only differed by 8.8 % from measured data.

8.2.2.2 *Carboxylic Acids*

Pure component saturated vapour pressure and density data for four short-chain carboxylic acids were calculated using the TraPPE-UA force field and the parameters developed specifically for carboxylic acids by Kamath et al. (2004). The studies of Kamath et al. (2004) were restricted to linear carboxylic acids, hence, simulation of acetic acid, propanoic acid and pentanoic acid was relatively straightforward. However, the force field as it stood was not capable of handling branched compounds such as 2-methyl propanoic acid, since the parameters for bond angles and torsions involving the -CH- functional group were not available. In this work the CH-C=O and CH-C-O angle energies are described using the same parameters that are used for CH_x-C=O and CH_x-C-O (x = 2 or 3) (Kamath, 2004). Parameters for the CH_3-CH-C=O, CH_3-CH-C-O and CH-C-O-H dihedral energies are assumed to be the same as those for the CH_x-CH_2-C=O, CH_x-CH_2-C-O and CH_x-C-O-H torsional energies. Hence, it is assumed that, in this instance, the number of hydrogens attached to the alpha carbon does not affect the force field. The torsional parameters involving the -CH_2- functional group were not listed in the original paper by Kamath et al. (2004) and were taken from existing torsions, viz.: O-C-O-H, CH_x-CH_2-CH_2-CH_x and CH_x-C-O-H, by Martin and Potoff (Martin, 2005). The TraPPE-UA parameters used in this project and those proposed here for the -CH- pseudo-atom are given in Tables B.1 to B.4.

The simulated results for the pure component acid systems are given in Table 8.15 and the saturated vapour pressures and vapour-liquid coexistence data are shown in Figures 8.19 to 8.22. The vapour pressures are presented as Clausius-Clapeyron plots to allow the critical pressure and normal boiling point to be determined (discussed below). The average statistical standard deviations calculated from the block averages are less than 1.25 % for the saturated liquid density and 8.5 % for the saturated vapour density and pressure. Previous simulation results are available for acetic acid and pentanoic acid (Kamath et al., 2004) and are included for comparison.

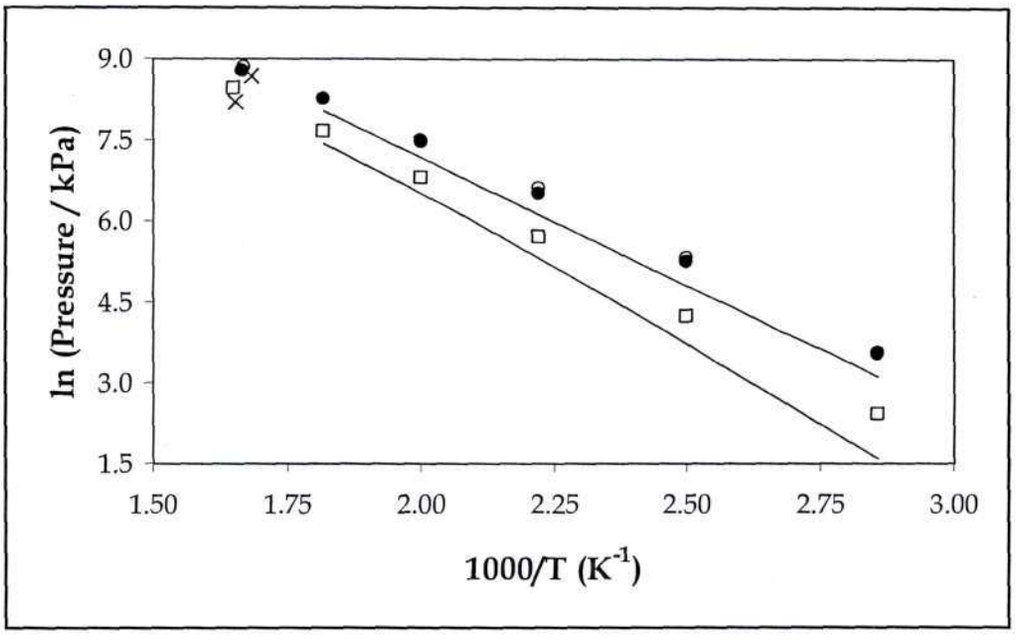

Figure 8.19: Clausius-Clapeyron plot for acetic acid (○) and 2-methyl propanoic acid (□). Experimental data are represented by solid lines and crosses (critical point) and symbols represent simulation results. Filled symbols are Kamath et al. (2004) data.

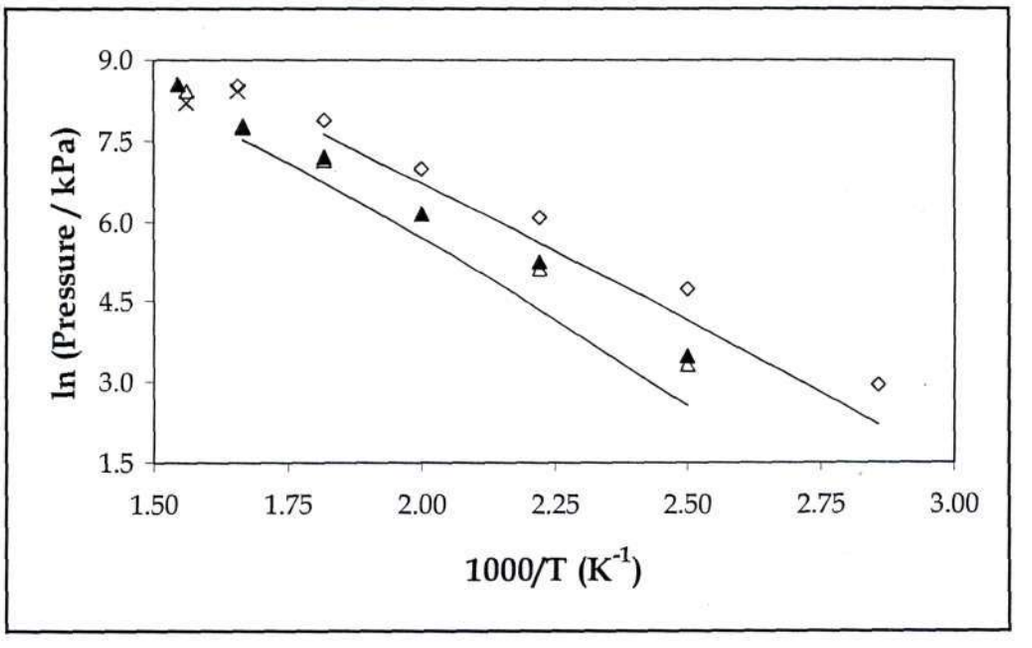

Figure 8.20: Clausius-Clapeyron plot for propanoic acid (◊) and pentanoic acid (Δ). Experimental data are represented by solid lines and crosses (critical point) and symbols represent simulation results. Filled symbols are Kamath et al. (2004) data.

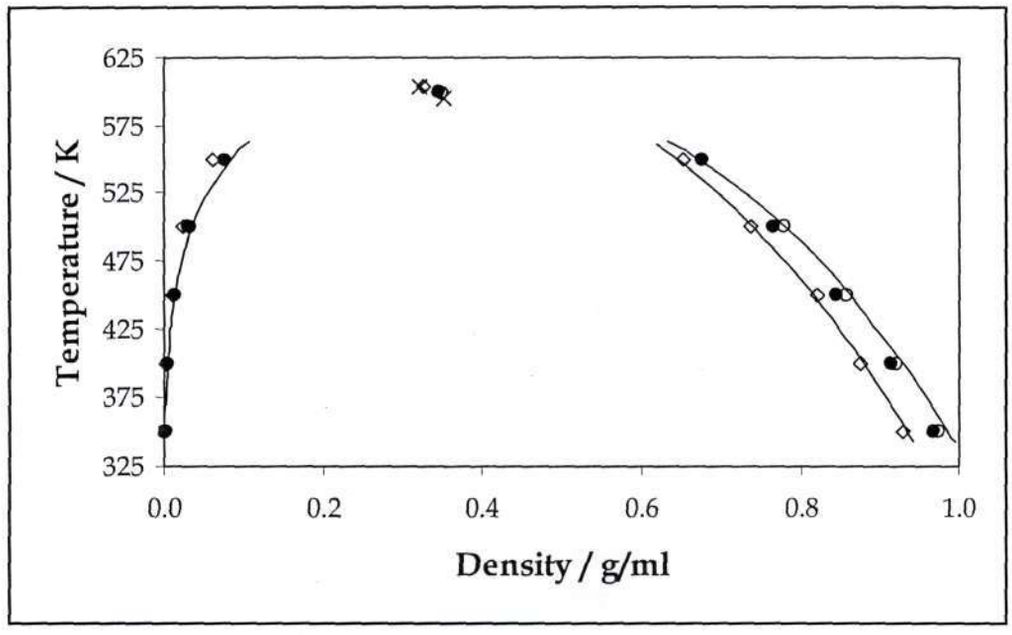

Figure 8.21: Vapour-liquid coexistence plot for acetic acid (○) and propanoic acid (◊). Experimental data are represented by solid lines and crosses (critical point) and symbols represent simulation results. Filled symbols are Kamath et al. (2004) data.

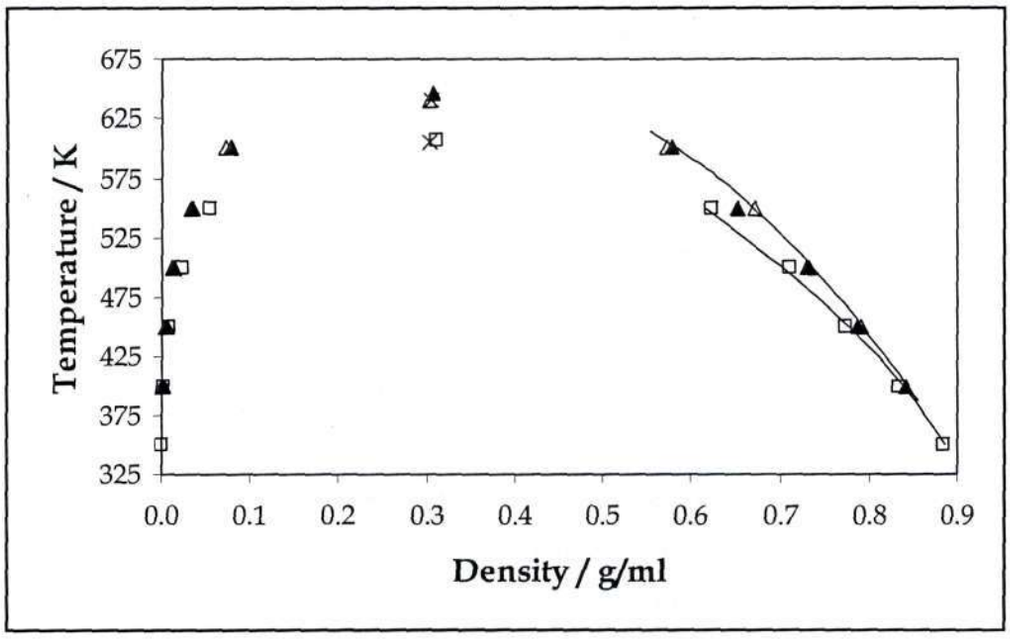

Figure 8.22: Vapour-liquid coexistence plot for 2-methyl propanoic acid (□) and pentanoic acid (Δ). Experimental data are represented by solid lines and crosses (critical point) and symbols represent simulation results. Filled symbols are Kamath et al. (2004) data.

Table 8.15: Numerical results for the pure component carboxylic acid simulations.

Compound	T (K)	P (kPa)	ρ_{liq} (g/ml)	ρ_{vap} (g/ml)
Acetic Acid	350	35.82	0.974	7.92E-04
	400	203.56	0.919	4.09E-03
	450	734.16	0.857	1.27E-02
	500	1744.83	0.778	3.10E-02
	550	3919.32	0.677	7.53E-02
Propanoic Acid	350	18.85	0.929	4.81E-04
	400	114.21	0.875	2.63E-03
	450	438.66	0.820	9.85E-03
	500	1092.60	0.736	2.34E-02
	550	2611.01	0.654	6.04E-02
2-methyl propanoic Acid	350	11.14	0.884	3.45E-04
	400	69.06	0.834	1.98E-03
	450	306.76	0.774	8.09E-03
	500	897.15	0.710	2.25E-02
	550	2110.33	0.622	5.47E-02
Pentanoic Acid	400	28.28	0.841	9.06E-04
	450	168.81	0.791	4.97E-03
	500	460.26	0.731	1.25E-02
	550	1260.40	0.671	3.50E-02
	600	2354.99	0.571	7.24E-02

The saturated vapour pressures (Figures 8.19 and 8.20) show that the TraPPE-UA force field predicts carboxylic acid vapour pressures that are larger than the experimental values, which is consistent with the results obtained by Kamath et al. (2004). This deviation from the experimental vapour pressures decreases with increasing temperature. According to Kamath and co-workers, this effect is directly related to the molecules aspect ratio $L^* = L/\sigma$, where L is the distance separating site centers and σ is the collision diameter (Kamath et al., 2004; Stoll et al., 2001). Although options do exist for improving the vapour pressure prediction through alteration of the aspect ratio (Kamath et al., 2004), the goal of the TraPPE-UA force field

parameterization was to accurately represent the saturated liquid densities and critical temperature, which effectively fixes the simulated vapour pressures (Kamath et al., 2004).

As is observed in Figures 8.21 and 8.22, the saturated liquid densities are all in close agreement with experiment (Daubert and Danner, 1989; DDB, 1998; Vargaftik, 1975), with average deviations of less than 1.1 %. Experimental vapour densities are only available for acetic acid (Vargaftik, 1975), and again the simulation results agree well with experiment (see Figure 8.21). For propanoic acid, 2-methyl propanoic acid and pentanoic acid, experimental saturated vapour densities are not available. However, the predicted critical densities (discussed below and shown in Figures 8.21 and 8.22) are all very close to the corresponding experimental values, indicating that the simulated vapour densities are also accurate.

The critical properties are presented in Table 8.16 (and are shown in Figures 8.19 to 8.22), along with the corresponding experimental data and (where applicable) the values obtained by Kamath et al. (2004). As for the alcohols, the density scaling law and law of rectilinear diameters (Rowlinson and Swinton, 1982) were used to fit the vapour-liquid coexistence data. During the fitting procedure, the effect of different values for the effective scaling exponent (β_{eff}) was investigated, constrained to reproduce the critical temperature and density. The optimal value was found to be 0.29, rather than the universal Ising value of 0.325. Interestingly, as was mentioned in Section 8.2.2.1, a similar result was found for alcohols by Chen et al. (2001). This would indicate that for polar compounds (whose coexistence curves differ in shape from those of non-polar species (Chen et al., 2001)), a lower value for the critical exponent provides a better representation of the vapour-liquid coexistence data. The predicted critical temperatures and densities are all within 1 % of experiment, except for 2-methyl propanoic acid where the density differed by approximately 2.5 % (however, the simulated critical temperature for 2-methyl propanoic acid is 606.30 K, showing very good agreement with the experimental value of 605.0 K).

The predicted normal boiling points of the four acids, listed in Table 8.16, all lie below the experimental values (this is directly related to the overestimation of vapour pressures by the TraPPE-UA model). However, the deviation from experiment was within 5 % in all cases. Since the TraPPE-UA model predicts vapour pressures in excess of the experimental values, it also over-predicts the critical pressures, as shown in Table 8.16.

As was noted above, in order to simulate 2-methyl propanoic acid using the TraPPE-UA force field certain assumptions regarding bond angle and torsion parameters were required. Figures 8.19 and 8.22 and Tables 8.15 and 8.16 reveal that the 2-methyl propanoic acid simulations produced results indistinguishable in terms of accuracy from the results for the linear acids parameterized by Kamath et al. (2004). This outcome demonstrates that the TraPPE-UA force field may be readily extended to include branched carboxylic acids.

Table 8.16: Critical properties and normal boiling points for the carboxylic acids.

Compound	Data Source	T_c (K)	ρ_c (g/ml)	P_c (kPa)	T_b (K)
Acetic Acid	Experiment	594.75	0.3512	5790	391.15
	Kamath et al.	600.26	0.3444	6574	379.80
	This work	598.90	0.3492	6801	378.57
Propanoic Acid	Experiment	604.00	0.3221	4529	414.30
	This work	604.07	0.3256	5013	397.79
2-methyl propanoic Acid	Experiment	605.00	0.3017	3698	427.60
	This work	606.30	0.3096	4761	412.09
Pentanoic Acid	Experiment	639.90	0.3040	3629	459.30
	Kamath et al.	646.40	0.3071	5230	435.70
	This work	640.44	0.3036	4612	437.77

8.2.2.3 *Compound Property Trends*

An interesting possibility afforded by the simulation of pure component data for a series of molecules of a particular class of compound (alcohols and carboxylic acids in this case) is to check whether force fields such as TraPPE-UA are capable of correctly reproducing compound property trends. Two properties will be focused on here, namely, the normal boiling point (T_b) and the critical temperature (T_c). The change observed in these properties with increasing carbon number is plotted in Figures 8.23 and 8.24 below, where the solid lines provide a guide to the eye regarding trends.

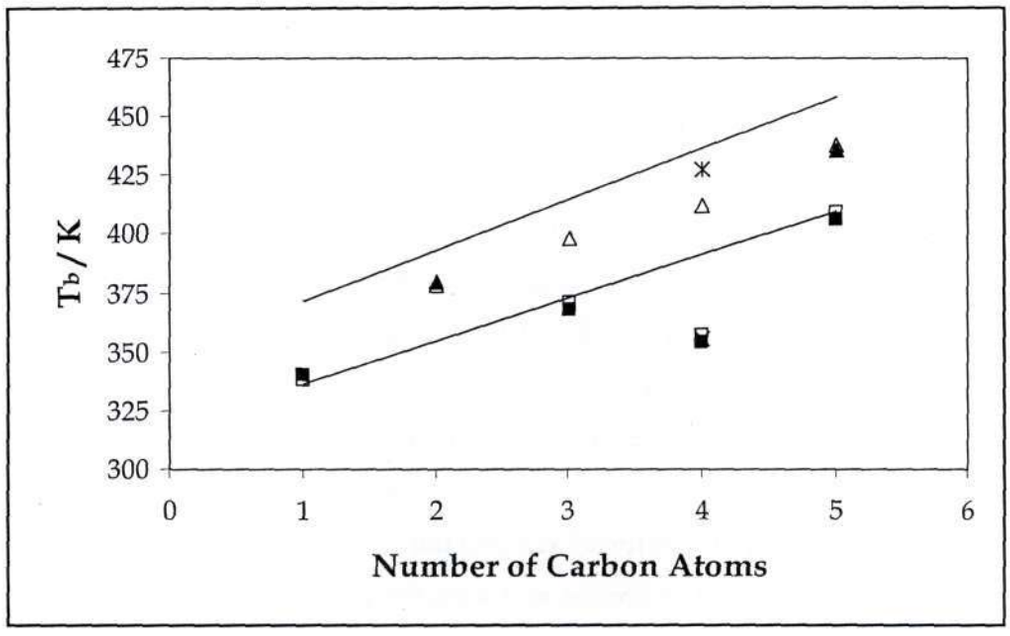

Figure 8.23: Plot of normal boiling point T_b versus number of carbon atoms for alcohols (□) and carboxylic acids (Δ). The solid lines represent experimental data and filled symbols are simulation results for alcohols (Chen et al., 2001) and carboxylic acids (Kamath et al., 2004).

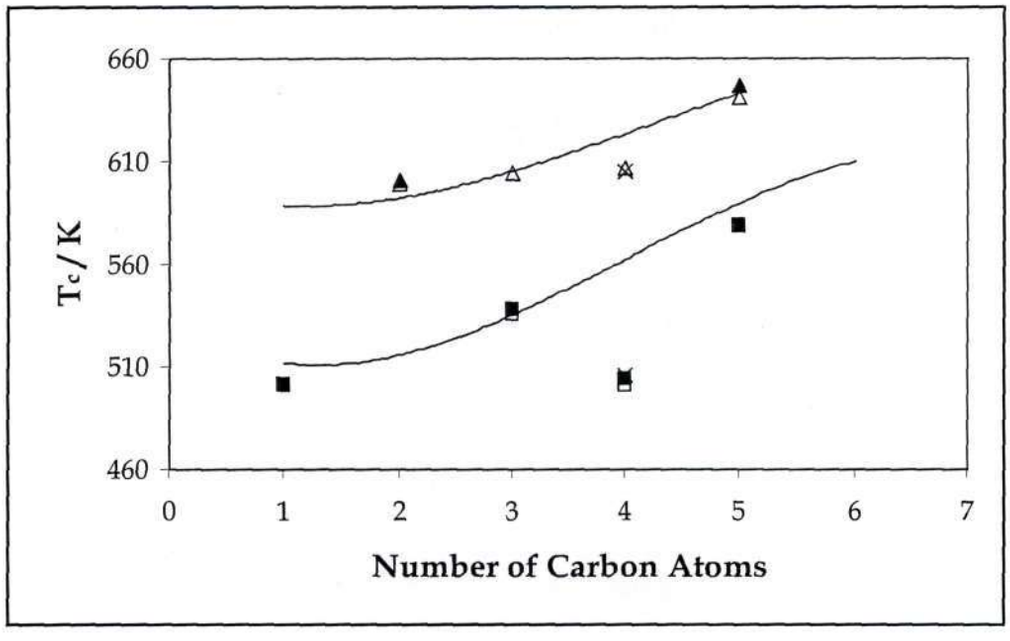

Figure 8.24: Plot of critical temperature T_c versus number of carbon atoms for alcohols (□) and carboxylic acids (Δ). The solid lines represent experimental data and filled symbols are simulation results for alcohols (Chen et al., 2001) and carboxylic acids (Kamath et al., 2004).

For the normal boiling point, both the alcohols and acids follow the trend exhibited by the experimental data well. The carboxylic acid T_b's are under-predicted, but this is a direct result of the vapour pressure over-prediction characteristic of the TraPPE-UA force field (discussed in Section 8.2.2.2). For critical temperature, the carboxylic acids follow the curved trend line of the experimental data remarkably well. The alcohols, on the other hand, appear to be following a straight-line trend instead and do not reproduce the curved critical temperature line clearly exhibited by the measured values. However, simulations for ethanol, 1-hexanol and 1-heptanol and perhaps even longer chain alcohols would be needed to draw a definitive conclusion.

It is a well-known fact that branched compounds have a lower boiling point and critical temperature than linear species of the same size and class. This effect is captured well by the TraPPE-UA force field, and the alcohol predictions are close to the experimental values (plotted as crosses and stars for the alcohols and the carboxylic acids, respectively, in Figures 8.23 and 8.24). However, it requires fairly close observation of Figures 8.23 and 8.24 to see the anticipated drop in normal boiling point (in particular) and critical temperature for the branched acid, 2-methyl propanoic acid. The extent of the boiling point depression is markedly less than that exhibited by the corresponding alcohol, 2-methyl-1-propanol and this dampening effect is attributed to the extensive dimerization known to occur in carboxylic acid systems. Although the simulation predictions for T_b are too low for all the acids, the degree to which the boiling point is depressed for 2-methyl propanoic acid is approximately 9 K, equivalent to that shown by the experimental data (see Figure 8.24).

8.2.3 Parameterization of H_2S, SO_2 and NO_2

The success of Monte Carlo molecular simulation relies, broadly, upon two principle foundations: the technology available in the form of computers and network hardware, and the potential energy models used to perform the actual calculations. It is the precision of these potential models (or force fields) that is essential for meaningful prediction of macroscopic properties, and the potential models rely completely on the parameters that form them for this accuracy. Poor parameterization will lead, however intricate or exact the potential model, to poor results. Thus, much care must be taken when parameterizing a force field and this process can often take months for a single

compound. Entire research groups devote themselves to the effort of producing new, efficient and precise force fields (and the associated parameters) which may be readily extended to a wide range of compounds. The Siepmann (TraPPE force field) and de Pablo (NERD force field) groups are good examples of research units seeking to provide high quality, transferable potentials that lack unnecessary complexity.

8.2.3.1 Parameterization Methodology

The potential energy model parameterized in this work takes the following form:

$$U(r_{ij}) = \sum_{nonbond} \left\{ 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi r_{ij}} \right\} + \frac{1}{2} \sum_{bonds} k_b (b - b_0)^2 + \frac{1}{2} \sum_{angles} k_\theta (\theta - \theta_0)^2 \tag{8.3}$$

where r_{ij} is the distance separating interaction sites i and j on different molecules and sites separated by more than four atoms on the same molecule, ε_{ij} and σ_{ij} are the Lennard-Jones energy and size parameters, respectively, q_i and q_j represent the atomic charges of sites i and j, k_b and k_θ are the harmonic constants for bond stretching and angle bending, respectively, b_0 is the equilibrium bond length and θ_0 is the equilibrium bond angle. For their simulations of H_2S, Kristóf and Liszi (1997) used a four-site model with three atomic sites and an additional point charge lying in the molecular plane on the bisector of the H-S-H angle (Kristóf and Liszi, 1997). Nath (2003) used the same potential model as Equation 8.3, with parameters for hydrogen adopted from the CFF force field (Hwang et al., 1998) and those for sulfur obtained by trial-and-error fitting to experimental pure component phase equilibrium data (Nath, 2003). The non-bonded terms between unlike interaction sites i and j were determined using the Lorentz-Berthelot mixing rules. This force field produced a good fit to experimental VLE data for H_2S and, hence, the potential function (Equation 8.3) was adopted for all three compounds parameterized in this work, with the unlike interaction parameters being explicitly determined.

The same basic parameterization methodology was followed for each of the three molecules. The Gaussian 98 suite of subroutines (Frisch et al., 1998) was used to obtain

the optimized (minimum energy) equilibrium bond lengths and angles and atomic charges (given as b_0, θ_0, q_i and q_j in Table 8.17, respectively). The calculations were carried out at the UMP2 level of theory using the 6-311+G(d,p) basis set, which yields equilibrium bond lengths and angles in agreement with experiment. Regions of the potential energy surfaces around the minimum energy bond lengths and angles were also determined using Gaussian 98. Harmonic force constants for bond stretching and angle bending, given in the last two terms of Equation 8.3, were then fitted to these energy curves and are given in Table 8.17 (i.e. contributions from the first term in Equation 8.3 were ignored at this stage). The average deviation between the energy calculated from the harmonic function and the UMP2/6-311+G(d,p) energy is less than 8.4 % for bond stretching and 3.1 % for angle bending for energies less than 4 kJ/mol (≈ 450 K, which are typical temperatures used in the simulations).

Table 8.17: Parameters for H_2S, SO_2 and NO_2[a]

Parameter	H_2S	SO_2	NO_2
q_i (e)	0.105	0.525	0.100
q_j (e)	-0.210	-0.2625	-0.050
b_0 (Å)	1.3334	1.4686	1.2028
θ_0 (deg)	92.082	119.258	133.926
k_b (K.Å^{-2})	158998.8	330944.8	978939.8
k_θ (K.rad^{-2})	28181.7	55202.6	62199.1
σ_{ii} (Å)	0.915	4.000	2.700
ε_{ii} (K)	1.900	30.0	115.0
σ_{ij} (Å)	3.350	2.125	3.880
ε_{ij} (K)	29.0	75.0	1.850
σ_{jj} (Å)	3.340	3.200	2.530
ε_{jj} (K)	355.0	133.5	305.0

[a] The subscript *i* refers to H, S and N for H_2S, SO_2 and NO_2, respectively.

Initial estimates for the Lennard-Jones parameters were determined from the change in UMP2/6-311+G(d,p) energy with intermolecular centre-of-mass separation between

two constrained and optimized molecules. Intermolecular energy curves were generated (using Gaussian 98) for different relative molecular orientations, and an example is given in Figure 8.25 for one of these orientations for two SO_2 molecules. Estimates for the Lennard-Jones 12-6 parameters were obtained by fitting Equation 8.3 to the potential curves for the different molecular orientations using a simple, iterative minimization code (Figure 8.25 also shows the fit to the ab initio data that was obtained). The parameters were then fine-tuned by fitting simulated equilibrium data to experimental results (Braker and Mossman, 1980; Daubert and Danner, 1989; Goodwin, 1983; Perry et al., 1997; Reamer and Sage, 1952; Vargaftik, 1975). The final parameters for H_2S, SO_2 and NO_2 are listed in Table 8.17.

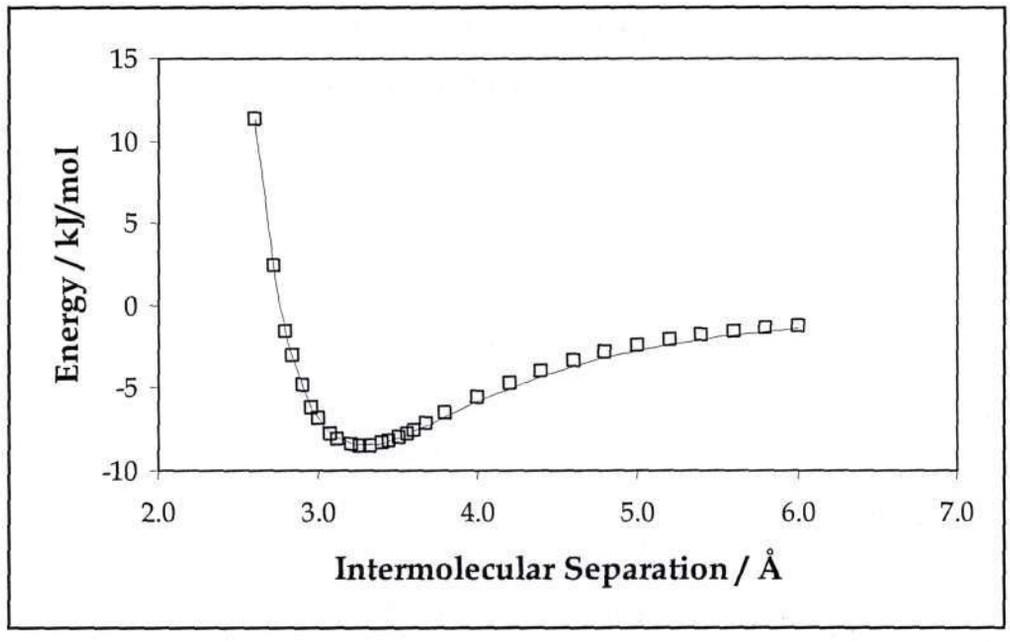

Figure 8.25: A typical potential energy curve generated by varying the intermolecular distance between two constrained and optimized SO_2 molecules. Symbols are the UMP2/6-311+G(d,p) energies generated by Gaussian 98 and the solid line is the fit obtained using Equation 8.3.

8.2.3.2 *Parameterization Results*

The simulation results for H_2S, SO_2 and NO_2 (using the final parameters given in Table 8.17) are presented in Table 8.18. The vapour-liquid coexistence curve and

saturated vapour pressure for H_2S are shown in Figures 8.26 and 8.27, together with the simulated data from the NERD potential model (Nath, 2003). The model developed in this project provides an equivalent level of accuracy to the NERD force field for the saturated liquid densities, and there is a significant improvement in the saturated vapour pressures and vapour densities. This is attributed to the increased flexibility acquired through the use of explicit parameters for unlike interactions.

Since the completion of the parameterization section of this project, Kamath et al. (2005) have published a three-site potential model for H_2S very similar to that proposed here (in fact, they presented four parameter sets with varying Lennard-Jones parameters and atomic charges). The potential model presented as part of this work and those of Kamath et al. (2005) appear to possess comparable predictive accuracy in terms of their fit to experimental phase equilibrium data, although it is difficult to come to a definite conclusion in this regard because Kamath et al. (2005) did not provide a table of simulation data that could be used for comparison. They did, however, plot their data, but the relatively small graphs with compressed x-axes make comparison between the model presented here and that of Kamath and co-workers uncertain. Undoubtedly, Kamath et al. (2005) were able to deliver a very accurate description of the pure component phase equilibrium data of H_2S and, in particular, the saturated vapour pressures. They acknowledge that this is due, at least in part, to the fact that they placed more importance on the reproduction of the saturated vapour pressures than the saturated liquid densities (Kamath et al., 2005). Consequently, for certain parameter sets (in particular, set A in the referred paper) they found that the saturated liquid densities showed significant deviation from experiment above 340 K (Kamath et al., 2005). In any event, it is encouraging that two different, independently-developed three-site force fields both produced an excellent description of the phase behaviour of a polar compound such as H_2S. This supports the conclusion of Kamath et al. (2005) that multiple parameter sets exist that are capable of reproducing both the pure component (and mixture) phase behaviour of a particular molecule.

An interesting aspect of the models of Kamath and co-workers is that parameter sets C and D, with heuristically selected partial charges, provided an excellent description of the pure component phase equilibrium data. In addition, the *P-x-y* description obtained for the H2S + pentane VLE system using sets C and D was superior to that of sets A and B (these sets had partial charges derived using the CHELPG methodology). This result indicates that, in some cases, a better description of mixture phase

behaviour may be attained by slight, heuristic (i.e. trial and error) adjustments to the atomic charges, without comprising the accuracy of the pure component predictions.

The reader's attention is drawn to the fact that the slightly more complicated four-site model of Kristóf and Liszi (1997) also provides an excellent description of the H_2S experimental phase equilibrium data. Other attempts to accurately model the phase behaviour of H2S have been attempted by Jorgensen (1986), Delhommelle et al. (1999) and Delhommelle et al. (2000).

Table 8.18: Simulation results for H_2S, SO_2 and NO_2 using the parameters given in Table 8.17.

H_2S			
T (K)	P (kPa)	ρ_v (g/ml)	ρ_l (g/ml)
225	202.71	0.0039	0.917
250	525.22	0.0093	0.873
275	1140.76	0.0189	0.825
300	2160.91	0.0360	0.772
325	3625.53	0.0591	0.705
340	4901.87	0.0832	0.654

SO_2			
T (K)	P (kPa)	ρ_v (g/ml)	ρ_l (g/ml)
275	171.51	0.0050	1.435
300	422.69	0.0116	1.362
325	882.83	0.0236	1.290
350	1714.97	0.0465	1.204
375	3021.78	0.0842	1.106

NO_2			
T (K)	P (kPa)	ρ_v (g/ml)	ρ_l (g/ml)
300	840.12	0.0165	1.435
325	1601.03	0.0303	1.362
350	2910.11	0.0560	1.289
375	4588.48	0.0910	1.196
400	7060.29	0.1438	1.074

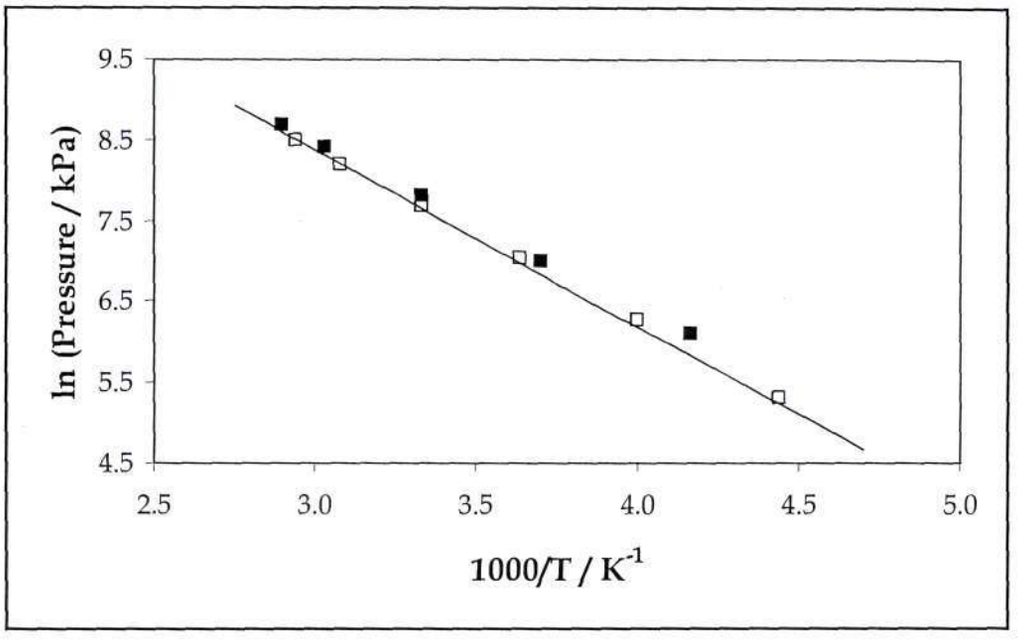

Figure 8.26: Saturated vapour pressure curve for H_2S (□). The solid line represents the experimental data and the symbols are the simulation data. Filled symbols indicate the data is that of Nath (2003).

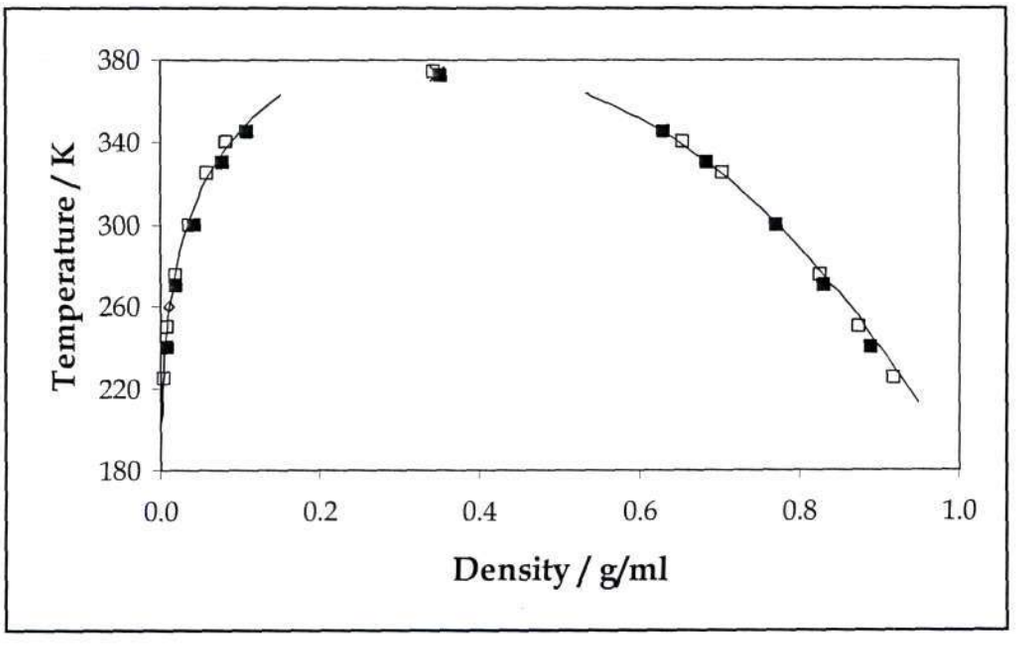

Figure 8.27: Vapour-liquid coexistence curve for H_2S (□). The solid line represents the experimental data and the symbols are the simulation data. Filled symbols indicate the data is that of Nath (2003) and the cross represents the experimental critical point.

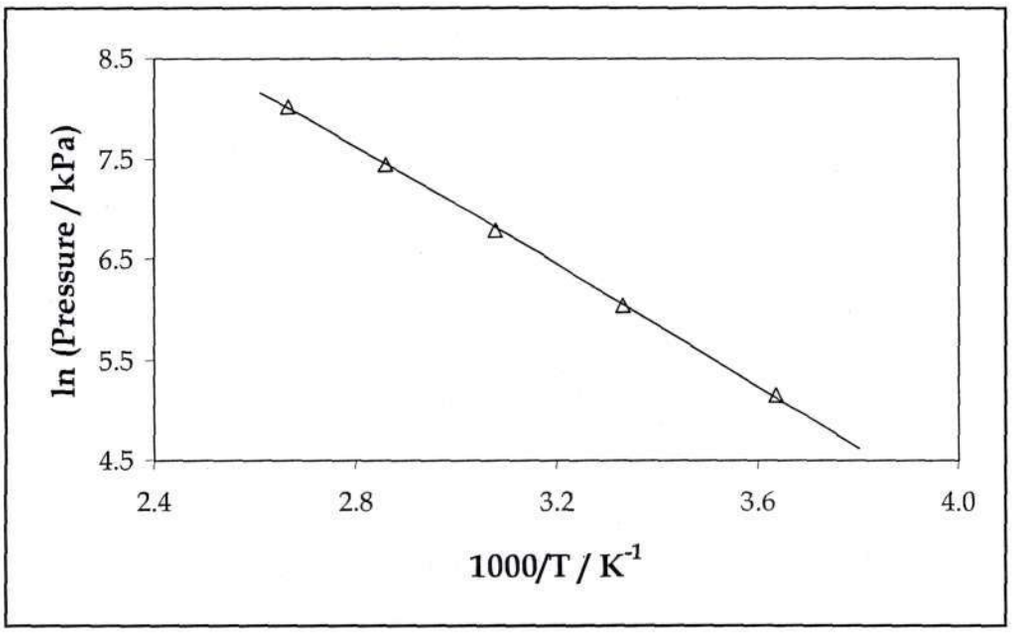

Figure 8.28: Saturated vapour pressure curve for SO_2 (Δ). The solid line represents the experimental data and the symbols are the simulation data.

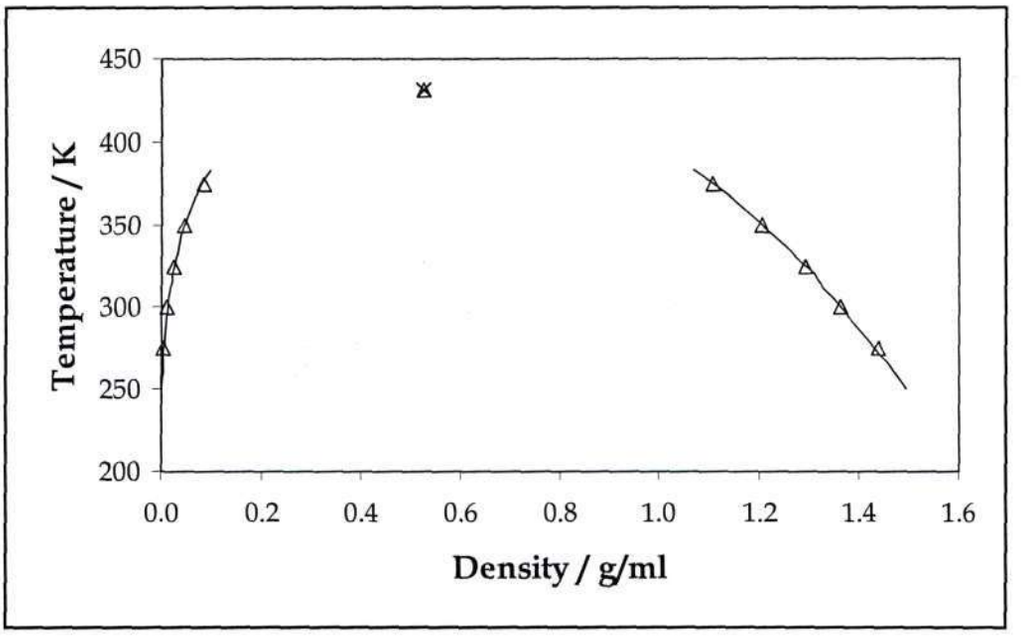

Figure 8.29: Vapour-liquid coexistence curve for SO_2 (Δ). The solid line represents the experimental data and the symbols are the simulation data. The cross represents the experimental critical point.

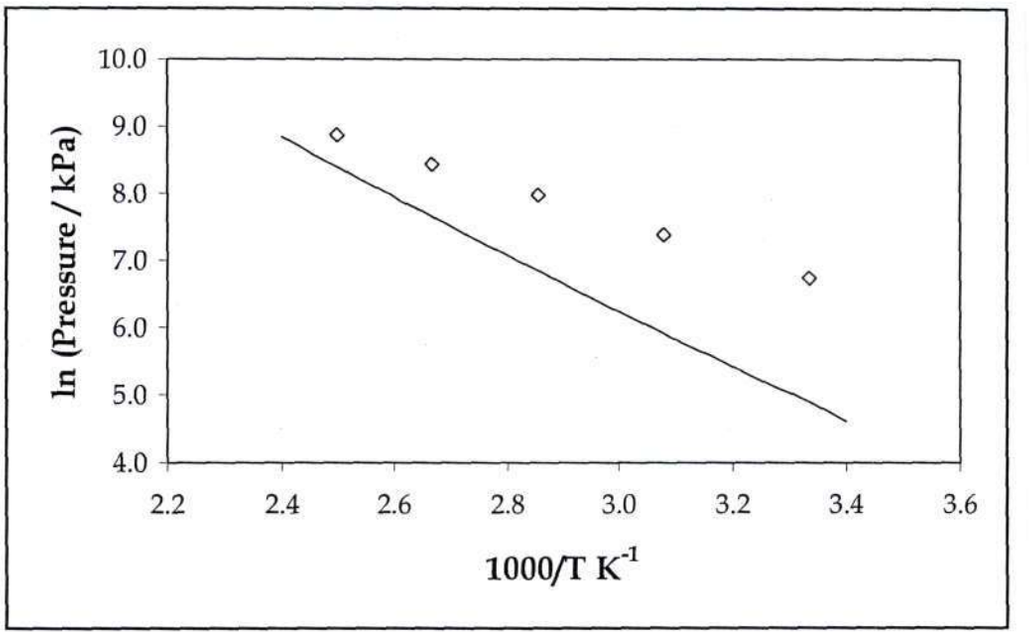

Figure 8.30: Saturated vapour pressure curve for NO_2 (◊). The solid line represents the experimental data and the symbols are the simulation data.

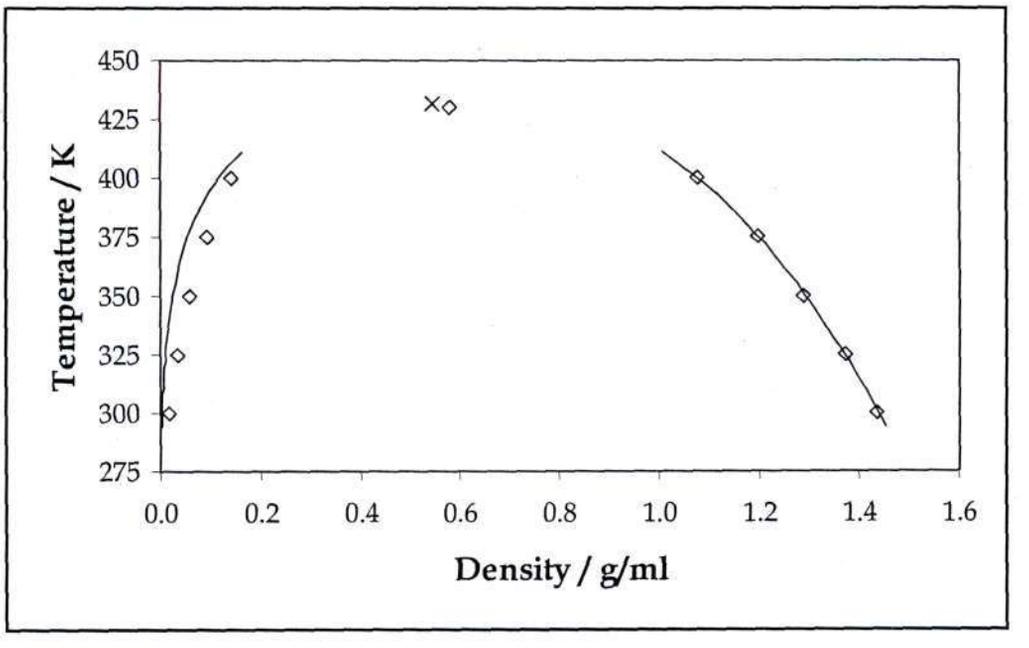

Figure 8.31: Vapour-liquid coexistence curve for NO_2 (◊). The solid line represents the experimental data and the symbols are the simulation data. The cross represents the experimental critical point.

The simulated saturated densities and vapour pressures for SO_2 are shown in Figures 8.28 and 8.29 and are in excellent agreement with the experimental data (Braker and Mossman, 1980; Perry et al., 1997; Vargaftik, 1975). This demonstrates that, in certain instances, even relatively simple potential models are capable of extremely accurate prediction of the phase behaviour of polar molecules.

Figures 8.30 and 8.31 illustrate that simulations based on the optimized parameters for NO_2 yield excellent saturated liquid densities, but despite extensive efforts to improve the fit, the corresponding saturated vapour densities and pressures are too large. Due to the lone electron on the NO_2 molecule, extensive dimerization occurs in the liquid phase and in the vapour phase at temperatures below 413.15 K (Cotton et al., 1999). In other words, individual molecules are rare and exist predominantly in the vapour phase and even then, only at relatively high temperatures. In fact, the liquid phase is almost completely dimerized and consists of more than 99 % N_2O_4 (Cotton et al., 1999). Evidently, the potential model developed here cannot satisfactorily describe this complex chemistry. The intention of the current work was to ascertain whether the relatively uncomplicated Lennard-Jones potential model was capable of successfully representing the phase behaviour of molecules like H_2S, SO_2 and NO_2. The results indicate that Lennard-Jones based force fields are quite capable of producing good data for certain polar compounds (e.g. SO_2 or H_2S), but that molecules that possess slightly more complicated chemistry like NO_2 become problematic. This shows that the current model and approach are insufficient and need some refinement. Further work is evidently required to produce an accurate model of the NO_2 phase behaviour. Areas of research that hold promise and should be explored are:

1. The introduction of Morse bonds (Morse, 1929) to simulate N_2O_4 dimerization. This may yield a better description of the observed NO_2 chemistry.
2. Alternatively, increasing the NO_2 partial charges could result in the expected dimerization behaviour. However, this would necessitate time consuming re-tuning of the Lennard-Jones parameters.
3. Another option is to develop a more intricate potential model (i.e. to move away from the Lennard-Jones form) that better represents the complex chemistry of NO_2.

Table 8.19: Lennard-Jones 12-6 parameters used for the plots in Figure 8.32 (the other parameters, e.g. atomic charges, were held constant and are given in Table 8.17).

Symbol	σ_{OO} (Å)	ε_{OO} (Å)	σ_{NO} (Å)	ε_{NO} (K)	σ_{NN} (K)	ε_{NN} (K)
□	2.750	325.0	4.100	1.750	2.500	100.0
+	2.500	300.0	3.900	1.850	2.750	125.0
–	2.500	300.0	3.900	1.850	2.250	125.0
○	2.650	320.0	4.000	1.850	2.600	135.0
◊	2.650	320.0	3.750	1.850	2.600	135.0
Δ	2.550	305.0	3.900	1.850	2.700	135.0
x	2.530	305.0	3.880	1.850	2.700	115.0

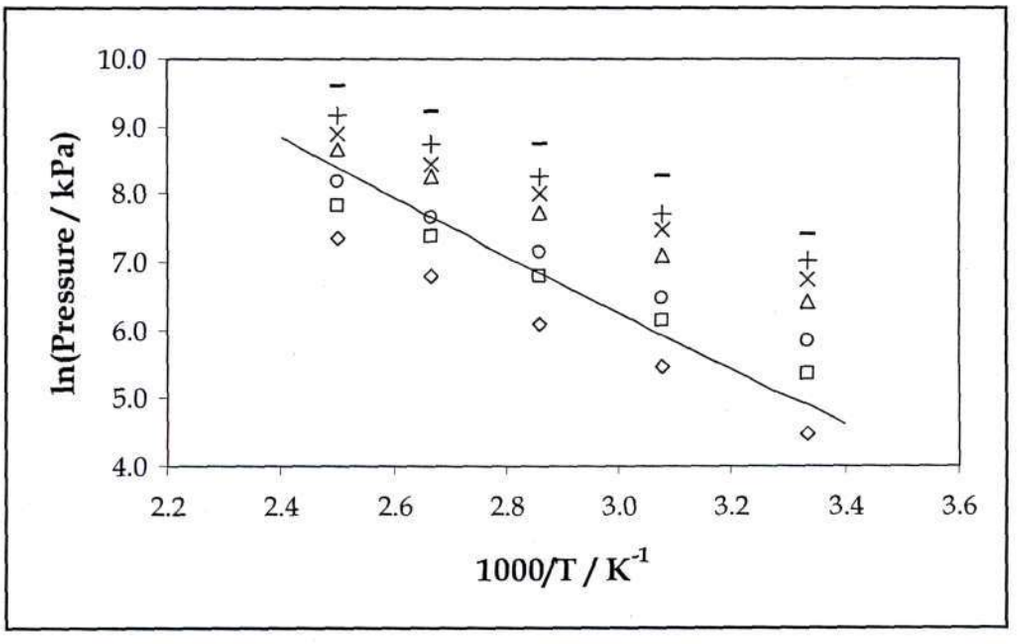

Figure 8.32: Plots of NO_2 saturated vapour pressures for the different parameter sets in Table 8.19 showing that the slope of the Clausius-Clapeyron plot is largely insensitive to the choice of intermolecular parameters.

It is difficult to obtain good agreement between simulated and experimental saturated vapour pressures with the model used here because of the different

qualitative dependence of the pressure on the temperature. This is shown in Figure 8.32, where the simulated saturated vapour pressures from several sets of intermolecular Lennard-Jones parameters (given in Table 8.19 above) are plotted against temperature. It is evident that altering the Lennard-Jones parameters merely shifts the saturated vapour pressure versus inverse temperature curve, and does not significantly affect its slope. Since the accuracy of the simulated saturated vapour densities correlate directly with that of the vapour pressures, this work focuses on correctly describing the saturated liquid densities. As was mentioned above, increasing the NO_2 atomic charges is one possible solution to this problem. Stoll et al. (2001) discuss how variation of the multipole moments can affect the vapour pressure curve.

The critical temperature and density of H_2S, SO_2 and NO_2 were estimated from the pure component vapour-liquid coexistence data by fitting to the density scaling law and law of rectilinear diameters (Allen and Tildesley, 1987). For alkane systems a scaling exponent (β_s) of 0.325 is usually used for the fitting. This exponent is commonly termed the universal Ising constant or critical exponent. However, for polar systems, certain authors (Chen et al., 2001; Nath, 2003) have found that a better fit to the coexistence density curves, critical temperatures and critical densities may often be obtained by allowing the scaling exponent to vary slightly. This effective scaling exponent, β_{eff}, offers an improved representation of the data because the coexistence curve differs in shape for polar species, limiting the range of applicability of the scaling law for polar compounds (Chen et al., 2001). The predicted critical temperatures and densities, together with the effective scaling exponent for each compound, are presented in Table 8.20. The simulated and experimental critical temperatures and densities are in good agreement for H_2S and SO_2, but for NO_2 the critical density exceeds the experimental data by 3.5 %. This is a direct result of the inflated vapour phase densities obtained from this model. In contrast, the critical temperature predicted for NO_2 differs by less than 1 K from the experimental value of 431.35 K (Braker and Mossman, 1980). The effective scaling factors that were used to obtain the critical data for H_2S and NO_2 are 0.315 and 0.290, respectively, which are somewhat lower than the Ising value of 0.325. Interestingly, for SO_2 a scaling exponent of 0.325 provided an excellent fit, indicating that the coexistence curve is similar in shape to that of rare gases and alkanes (Chen et al., 2001).

Table 8.20: Predicted critical temperatures, critical densities and scaling exponents.

	H_2S		SO_2		NO_2	
	Predicted	Experiment	Predicted	Experiment	Predicted	Experiment
T_c (K)	373.99	373.40	430.94	430.75	430.53	431.35
ρ_c (kPa)	0.342	0.348	0.524	0.525	0.577	0.557
β_{eff}	0.315		0.325		0.290	

8.2.4 Binary Phase Equilibrium Results

This section is devoted to the discussion of the simulation predictions for various binary vapour-liquid equilibrium (VLE) systems. The first set of results that will be presented are for systems consisting only of carboxylic acids, with the acid systems chosen to coincide with those modeled using the chemical theory (Section 8.1). To this end, four of the eight systems available were simulated: an isobar at 20 kPa and an isotherm at 393.15 K for both the propanoic acid + pentanoic acid and 2-methyl propanoic acid + pentanoic acid systems.

Thereafter, the new force field parameters developed in Section 8.2.3 for H_2S and SO_2 were used in a series of two-component VLE simulations to ascertain their ability to accurately predict high-pressure phase equilibria. H_2S and SO_2 have significant industrial importance because they occur as troublesome pollutants and byproducts of various petrochemical and refining processes (Speight, 1998). Their environmental impact has been well documented (Kroschwitz et al., 1997; Landsberg, 1981; Smith, 1974) and they are acknowledged as extremely hazardous chemicals, having a profound effect on the human respiratory system and reducing the body's resistance to lung infections. This results in susceptibility to diseases such as bronchitis and other lung ailments (Bates, 1972; Fairchild et al., 1972; Goldstein et al., 1973; Lisk, 1988; Pearlman et al., 1971). It is, therefore, imperative that they are removed by, for example, the Claus, rectisol or Stretford processes (Kohl and Riesenfeld, 1979; Speight, 1998). These operations strip the dangerous sulfur compounds using light alkanes and alcohols and, hence, the systems simulated here were selected for study.

For H_2S, the second components were: methane, ethane, propane and methanol, and for SO_2: methane, propane and ethanol. A general scarcity of data for VLE systems involving H_2S and SO_2 with short-chain alkanes and alcohols emphasized the need for a convenient and, most importantly, accurate method other than laboratory measurement that can be used to obtain this data.

8.2.4.1 Carboxylic Acid Binary VLE Results

When setting up the input files for these binary simulations, it is important to select initial volumes and numbers of molecules that do not substantially differ from the final equilibrium values (McKnight et al., 2005). However, it is also crucial that one does not initialize the system at a pre-supposed equilibrium and begin analysis before one has reached the true equilibrium. In addition, using Raoult's law to estimate the starting partial pressures, as suggested by McKnight and co-workers (2005), cannot be used for the strongly associating, non-ideal carboxylic acids studied here. Therefore, in this work experimental data (Clifford et al., 2005) was used to estimate reasonable initial pressures. The combined volume of the two simulation boxes was adjusted so that approximately two-thirds of the molecules were in the liquid box. It is important to note that a substantial change in average VLE properties was observed in the simulations before data was collected, showing that the systems had equilibrated and that the results do not pertain to the initial conditions.

The results presented here represent the first time that binary VLE data have been predicted for carboxylic acid systems using the Monte Carlo simulation technique. The TraPPE-UA force field has been used with the *NPT* Gibbs ensemble to simulate (as was mentioned above) an isobar at 20 kPa and an isotherm at 393.15 K for each of the following systems: propanoic acid + pentanoic acid and 2-methyl propanoic acid + pentanoic acid. The simulation results are plotted in Figures 8.33 to 8.40, with the experimental data of Clifford et al. (2005) being used for comparison.

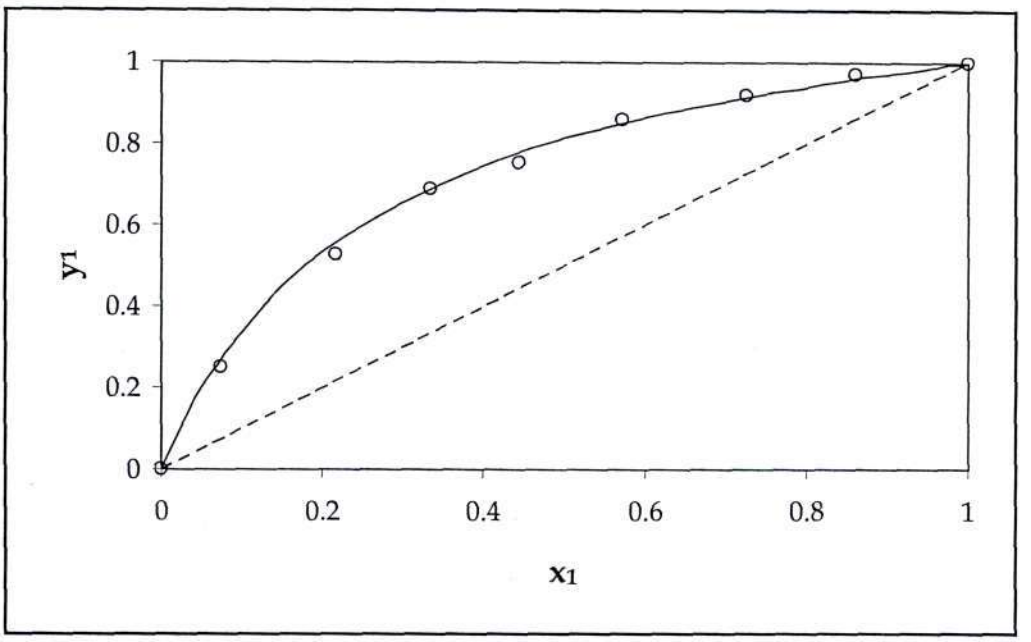

Figure 8.33: Plot of *x-y* data for propanoic acid (1) + pentanoic acid (2) at 393.15 K. The solid line represents the experimental data and the symbols the simulation results.

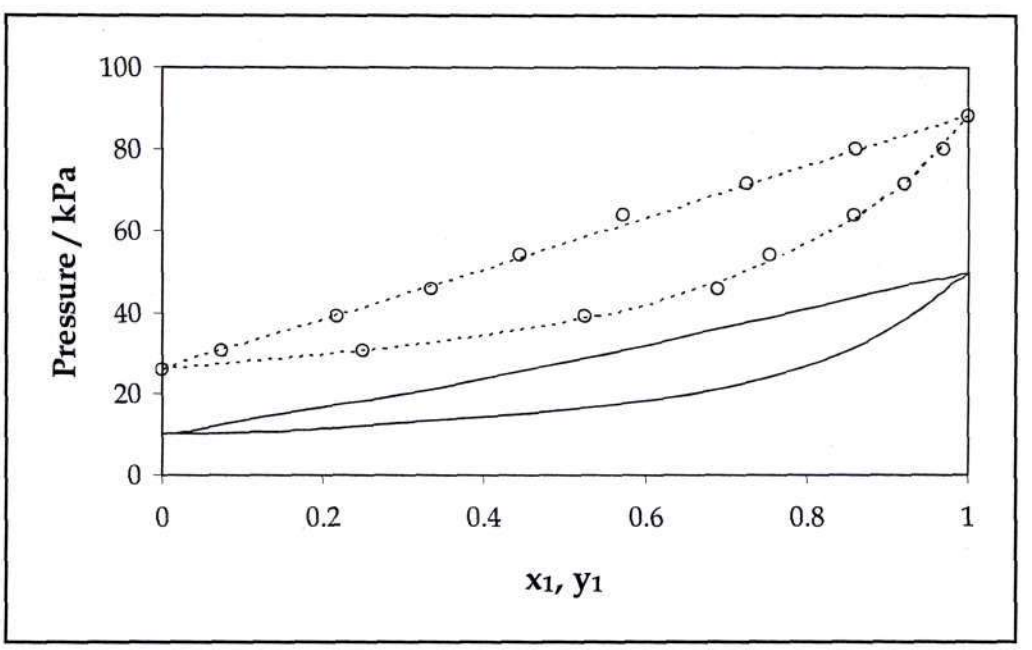

Figure 8.34: Plot of *P-x-y* data for propanoic acid (1) + pentanoic acid (2) at 393.15 K. The solid line represents the experimental data and the symbols the simulation results. The dashed line through the simulation data is a guide for the eye.

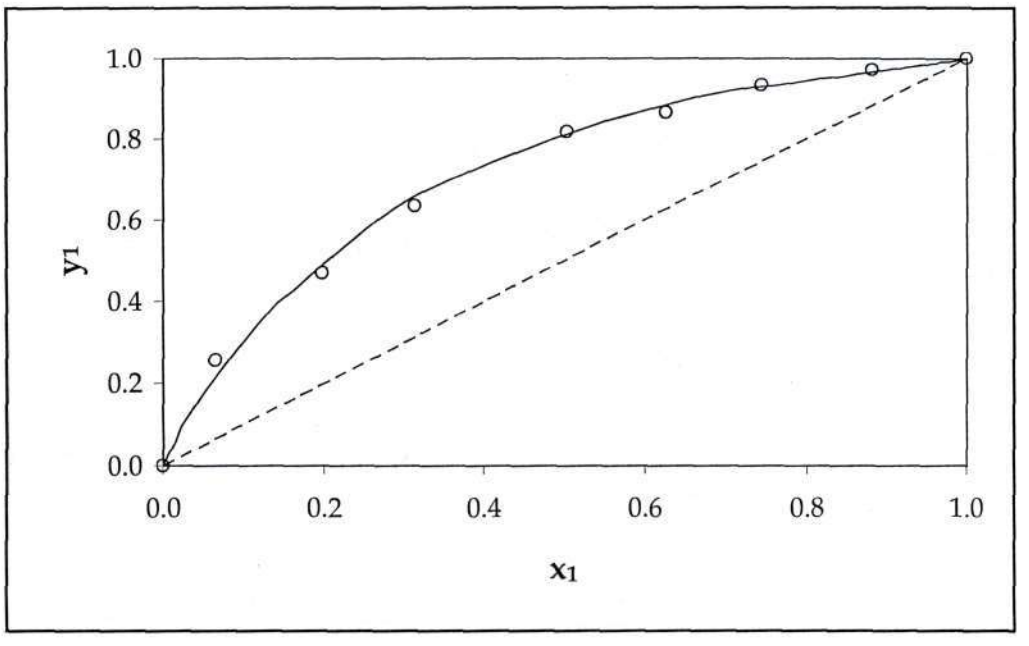

Figure 8.35: Plot of *x-y* data for propanoic acid (1) + pentanoic acid (2) at 20 kPa. The solid line represents the experimental data and the symbols the simulation results.

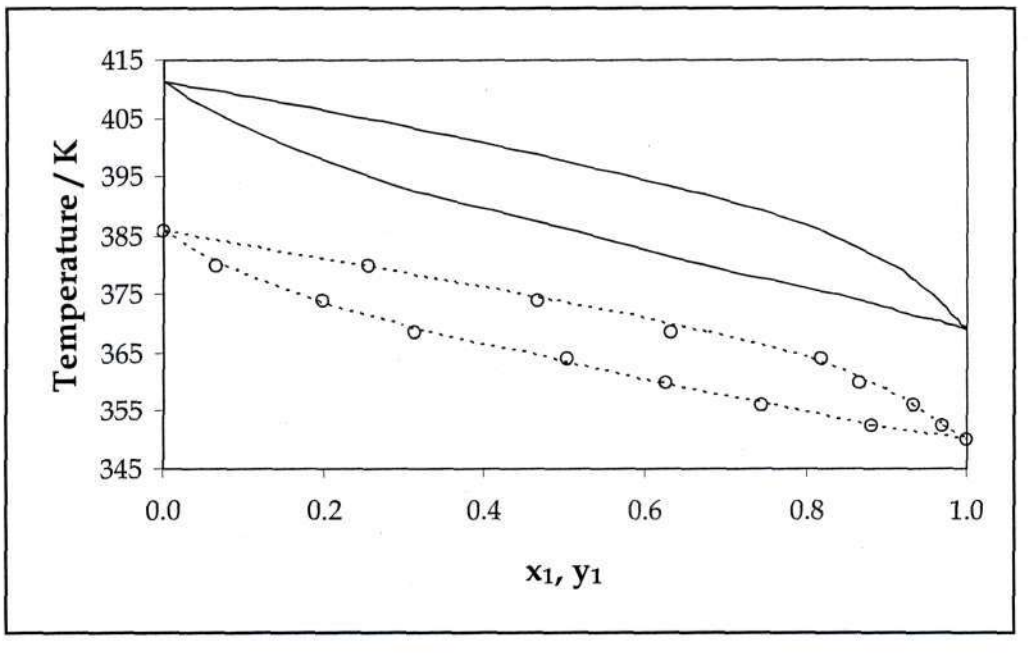

Figure 8.36: Plot of *T-x-y* data for propanoic acid (1) + pentanoic acid (2) at 20 kPa. The solid line represents the experimental data and the symbols the simulation results. The dashed line through the simulation data is a guide for the eye.

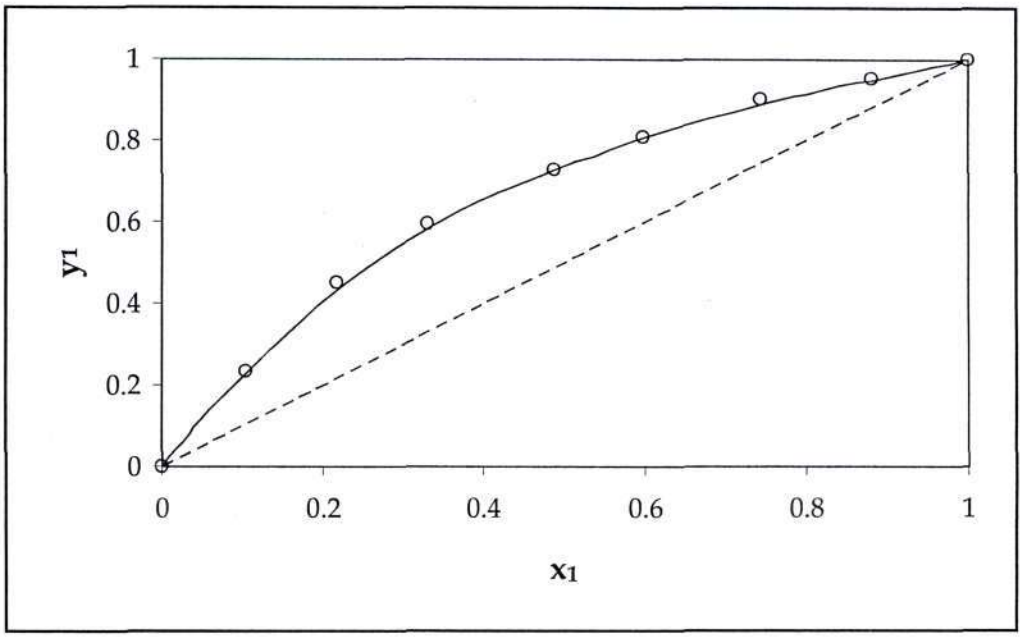

Figure 8.37: Plot of *x-y* data for 2-methyl propanoic acid (1) + pentanoic acid (2) at 393.15 K. The solid line represents the experimental data and the symbols the simulation results.

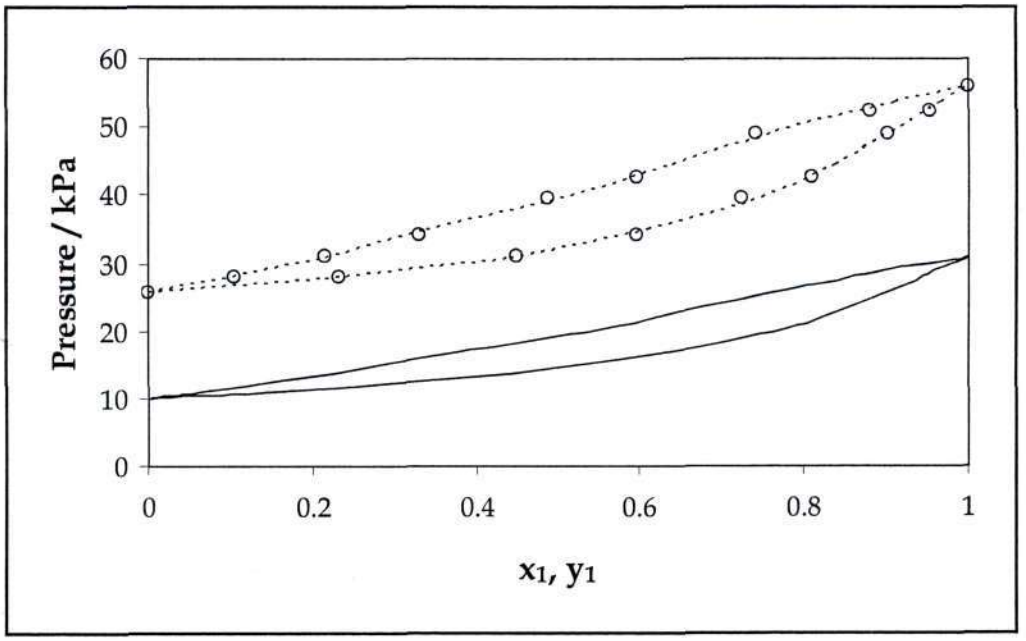

Figure 8.38: Plot of *P-x-y* data for 2-methyl propanoic acid (1) + pentanoic acid (2) at 393.15 K. The solid line represents the experimental data and the symbols the simulation results. The dashed line through the simulation data is a guide for the eye.

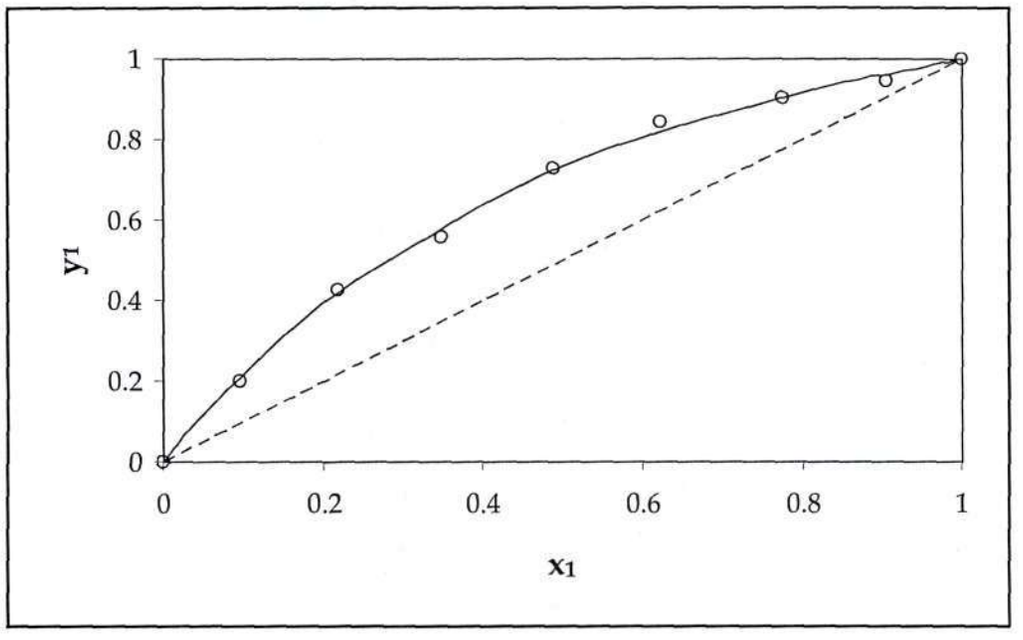

Figure 8.39: Plot of *x-y* data for 2-methyl propanoic acid (1) + pentanoic acid (2) at 20 kPa. The solid line represents the experimental data and the symbols the simulation results.

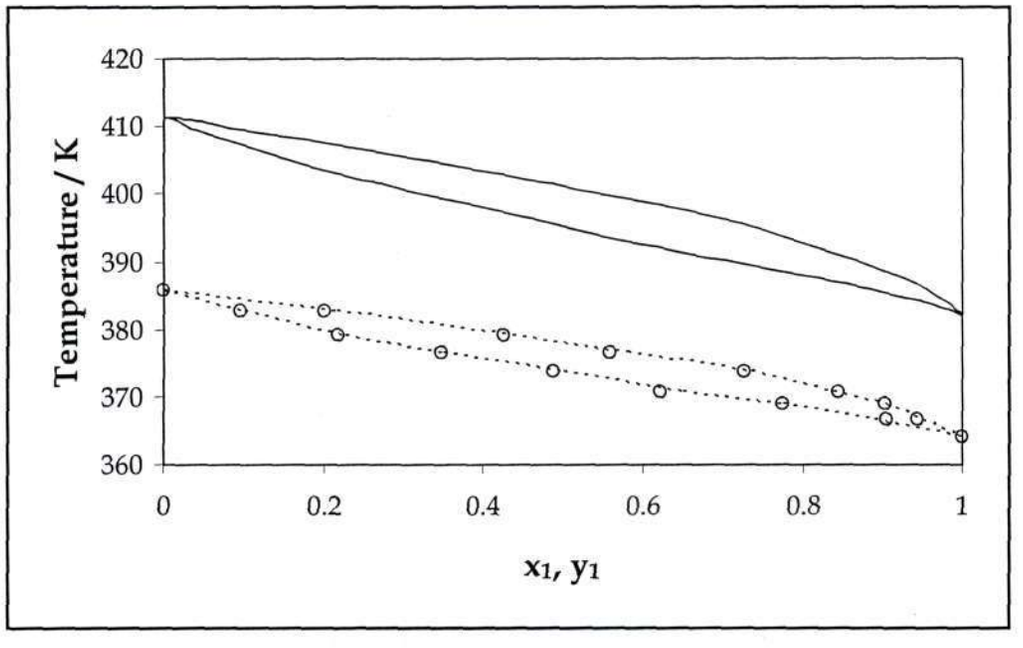

Figure 8.40: Plot of *T-x-y* data for 2-methyl propanoic acid (1) + pentanoic acid (2) at 20 kPa. The solid line represents the experimental data and the symbols the simulation results. The dashed line through the simulation data is a guide for the eye.

For both systems, the simulated composition (x-y) diagrams are in close agreement with the measured experimental data and the P-x-y phase diagrams (Figures 8.34 and 8.38) and T-x-y diagrams (Figures 8.36 and 8.40) show the anticipated pressure over-prediction and temperature under-prediction which come about as a direct consequence of the TraPPE-UA vapor pressure overestimation for carboxylic acids discussed in Section 8.2.2.2. These deviations are within the expected range of 15 - 40 kPa and 20 - 30 K (for the isobars and isotherms respectively), based on the departure from experimental data observed for the pure component simulations. One of the consequences of this effect is the slight widening of the P-x-y phase envelopes and narrowing of the T-x-y phase envelopes evident in the simulation results.

The end-points plotted in Figure 8.33 to 8.40 are taken directly from the pure component simulation results for propanoic, 2-methyl propanoic and pentanoic acid. These particular carboxylic acids were selected (along with acetic acid that had been previously simulated in the literature) for the pure component simulation portion of the project with the binary VLE systems in mind.

8.2.4.2 Hydrogen Sulfide Binary VLE Results

The simulation results for the H_2S binary systems are, in general, very good and reproduce the experimental data well, as is illustrated in Figures 8.41 to 8.48. The experimental data used for comparison are taken from several sources: for the H_2S + methane system, Reamer et al. (1951); for H_2S + ethane and H_2S + propane, Gmehling and Onken (1982) and for H_2S + methanol, Leu et al. (1992).

Following the pure component simulations, it was anticipated that the binary system pressures would be slightly over-predicted. However, it is evident from the graphs plotted below that this was not the case and, for the most part, the system pressures are accurately calculated. The exception is the H_2S + methanol system, where it is clear from Figure 8.48 that the simulated system pressures are too high. This effect visibly decreases as the H_2S concentration increases and, hence, the over-estimation is largely attributed to the TraPPE-UA force field over-predicting the alcohol partial pressure. The fact that the compositions (and, in particular, the liquid compositions) are under-predicted to some extent for the H_2S + methanol system (i.e. the solubility of H_2S in methanol is under-predicted) exacerbates this effect somewhat.

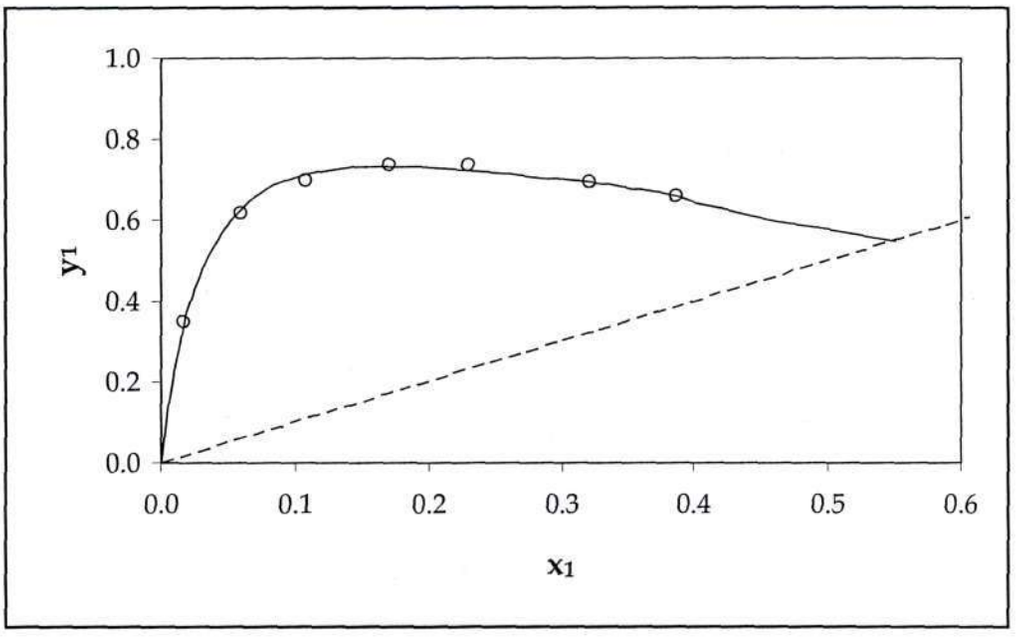

Figure 8.41: Plot of *x*-*y* data for methane (1) + hydrogen sulfide (2) at 277.59 K. The solid line represents the experimental data and the symbols the simulation results.

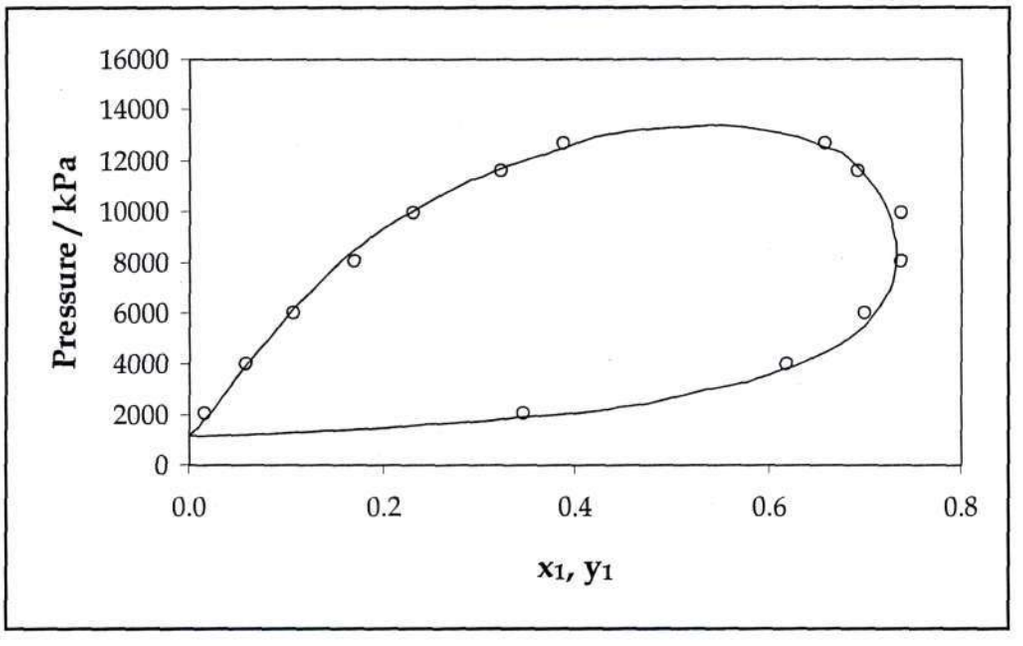

Figure 8.42: Plot of *P*-*x*-*y* data for methane (1) + hydrogen sulfide (2) at 277.59 K. The solid line represents the experimental data and the symbols the simulation results.

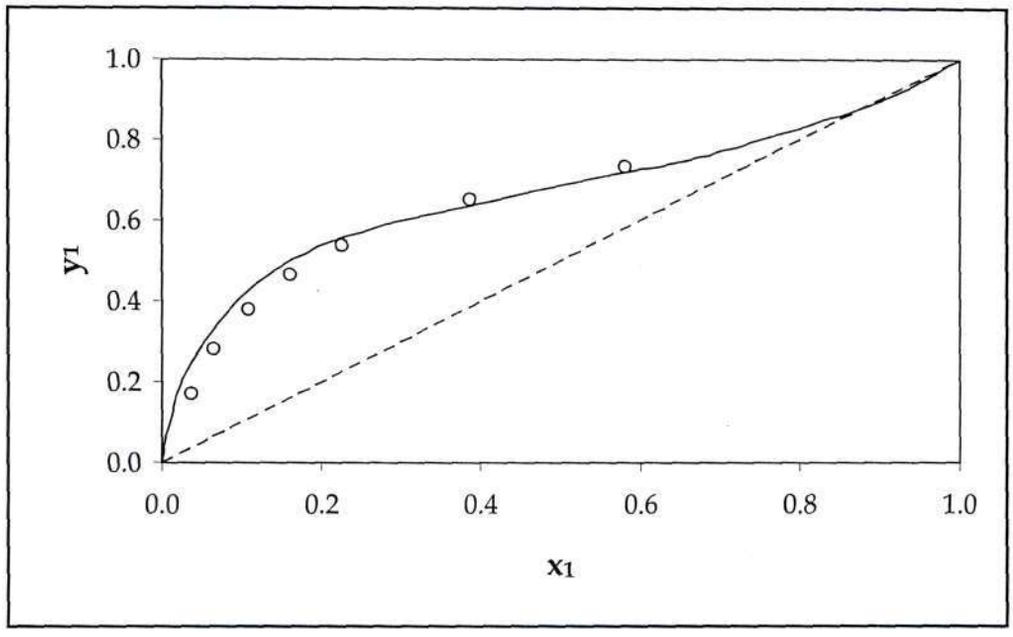

Figure 8.43: Plot of *x-y* data for ethane (1) + hydrogen sulfide (2) at 255.32 K. The solid line represents the experimental data and the symbols the simulation results.

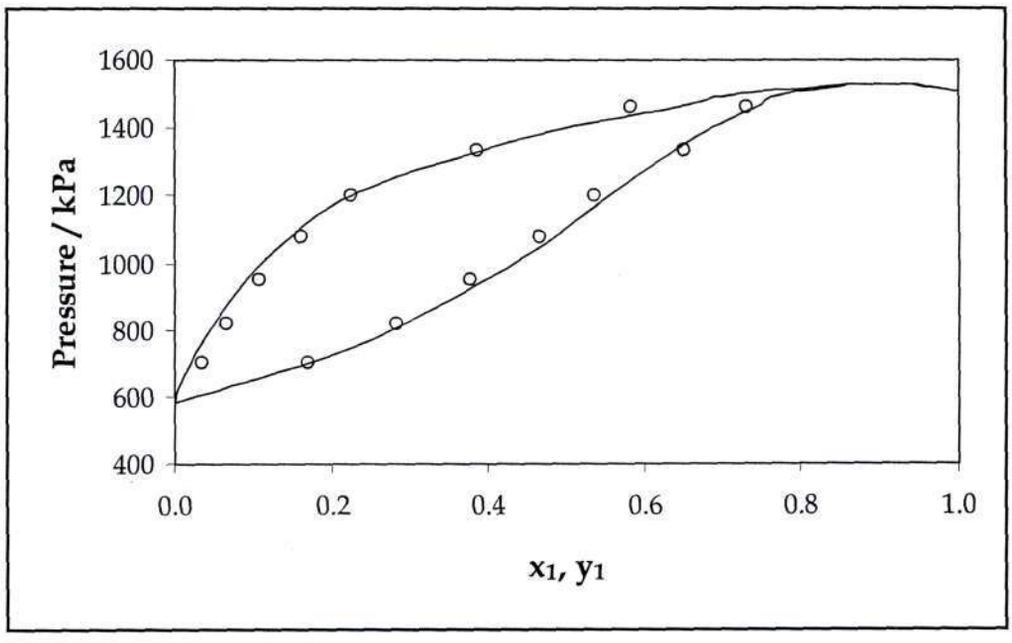

Figure 8.44: Plot of *P-x-y* data for ethane (1) + hydrogen sulfide (2) at 255.32 K. The solid line represents the experimental data and the symbols the simulation results.

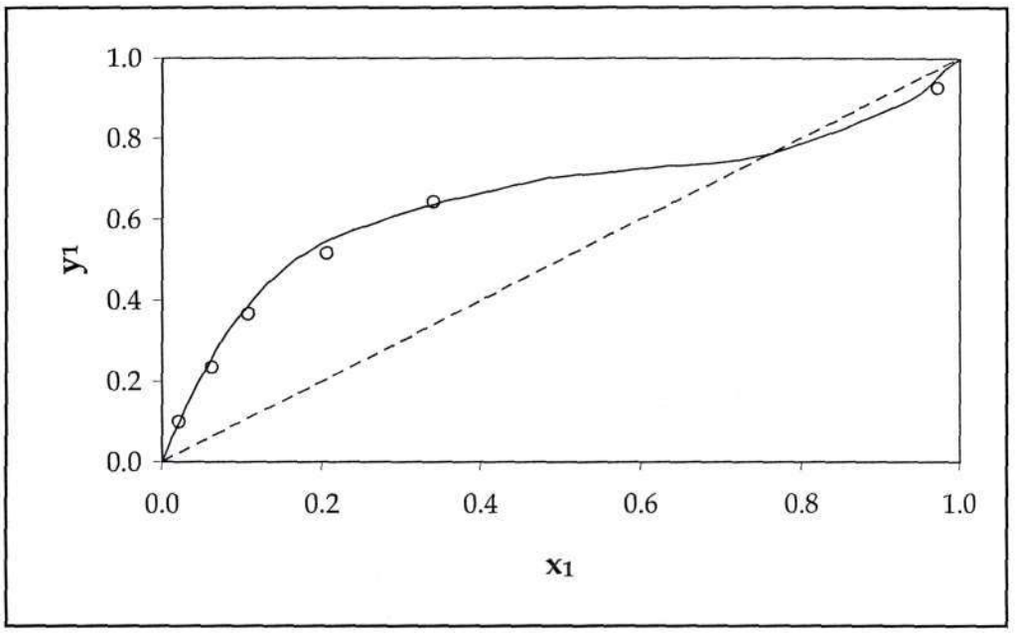

Figure 8.45: Plot of *x-y* data for hydrogen sulfide (1) + propane (2) at 344.7 kPa. The solid line represents the experimental data and the symbols the simulation results.

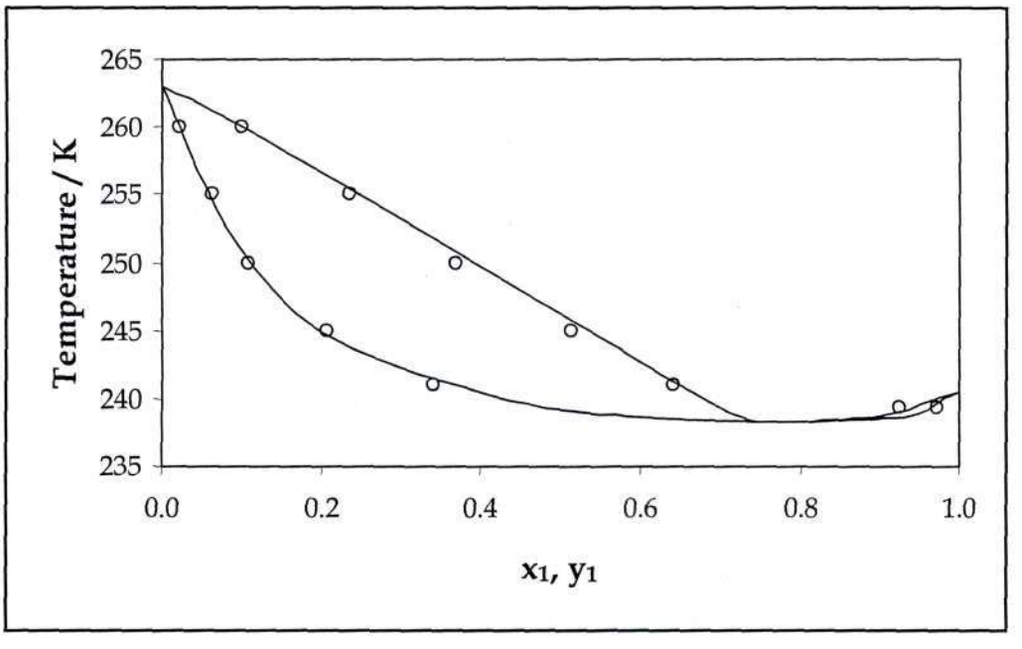

Figure 8.46: Plot of *T-x-y* data for hydrogen sulfide (1) + propane (2) at 344.7 kPa. The solid line represents the experimental data and the symbols the simulation results.

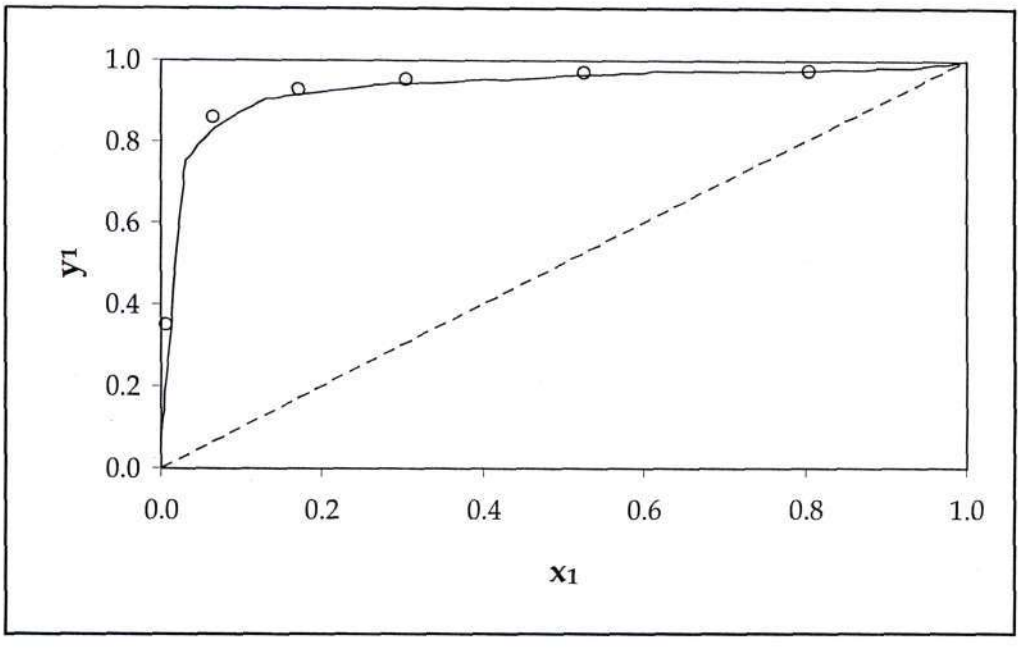

Figure 8.47: Plot of *x-y* data for hydrogen sulfide (1) + methanol (2) at 348.15 K. The solid line represents the experimental data and the symbols the simulation results.

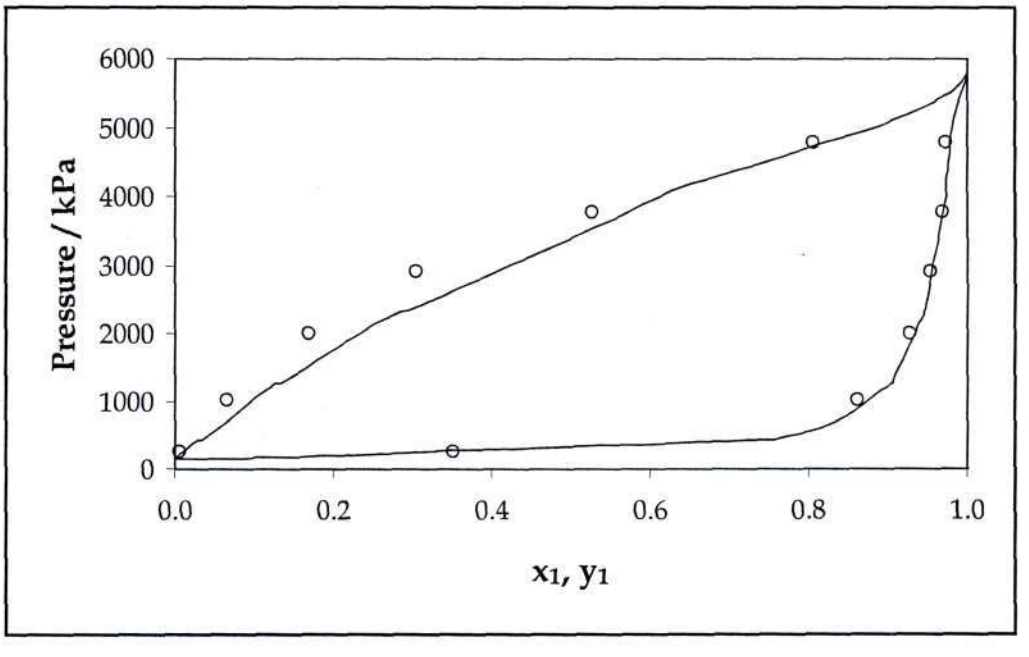

Figure 8.48: Plot of *P-x-y* data for hydrogen sulfide (1) + methanol (2) at 348.15 K. The solid line represents the experimental data and the symbols the simulation results.

For the methane (Figures 8.41 and 8.42) and propane (Figures 8.45 and 8.46) systems the simulated data are in good agreement with the experimental data, and the x-y, the P-x-y and the T-x-y plots (respectively) are accurately predicted, including the azeotropes. However, for the ethane system, the liquid mole fractions are over-predicted at first, but this trend is reversed at the highest simulated pressure. Conversely, the vapour mole fractions are generally under-predicted, as can be easily observed in Figures 8.43 and 8.44. For this reason, the azeotropic composition is under-predicted slightly, whilst the azeotropic pressure (± 1530 kPa) is accurately predicted.

An interesting aspect of the H_2S + methane simulations is that for the system temperature of 277.59 K, methane is supercritical. The agreement with experiment achieved for this system indicates that a force field (in this case, TraPPE-UA) that was parameterized by fitting to subcritical properties such as the vapour-liquid coexistence curve remains effective and accurate in the supercritical region. A similar result was obtained by Martin et al. (2000) when simulating an ethane + n-heptane temperature-composition phase envelope at high pressure (5.5 MPa).

The mixing rule of Halgren (1992) gave the best simulation results for the H_2S binary VLE systems, except for the ethane system where a combined mixing rule was used. This is discussed in greater detail in Section 8.2.5.2 below, where the sensitivity of the binary simulation results to different combining rules is illustrated.

8.2.4.3 Sulfur Dioxide Binary VLE Results

Overall, the results for the binary VLE systems involving SO_2 are not as good as those for H_2S. This outcome was rather unexpected since the pure component results show exactly the opposite trend. A possible reason is that the mixing rules used in this project are not entirely suited to systems involving SO_2, and improved results could be achieved through the development of a new combining rule aimed specifically at such systems. Time constraints did not allow this option to be explored. Certainly, the introduction of fitted, system-specific binary parameters into the combining rules (similar to those used in the Wong-Sandler (1992) and Twu-Coon (1996) mixing rules) would have significantly increased simulation accuracy. However, this would directly oppose one of the principle aims of molecular simulation, which is to provide accurate phase equilibrium data predictions without the requisite experimental data.

The results for the methane + SO_2 system at 301.48 K are shown in Figures 8.49 and 8.50 and are compared to the experimental data of Ohe (1990). Again, at this temperature methane is supercritical, adding complexity to the selection of the simulation starting point. The liquid mole fractions are over-predicted at all pressures, with this effect worsening steadily as the pressure rises (clearly illustrated in the *x-y* plot, Figure 8.49). The vapour mole fractions are fairly accurate at low pressure, but as the system pressure increases the SO_2 concentration is progressively under-predicted.

It is clear from Figure 8.50 that the azeotropic pressure is significantly under-predicted, with the simulation results indicating a pressure of approximately 25 MPa. The experimental data did not extend beyond 30 MPa, but a rough extrapolation from the available data points to an azeotropic pressure around 60 MPa. These extremely high pressures are a direct result of the supercritical behaviour of methane, shifting the simulations into a region well above the saturated vapour pressures used during parameterization of SO_2 and methane. It is speculated that the force fields used here (fitted to low pressure data) struggle to contend with the enormous system pressures. Simulations at pressures of 25 MPa and above failed to separate into discrete phases, with the liquid and vapour phases indistinguishable from one another.

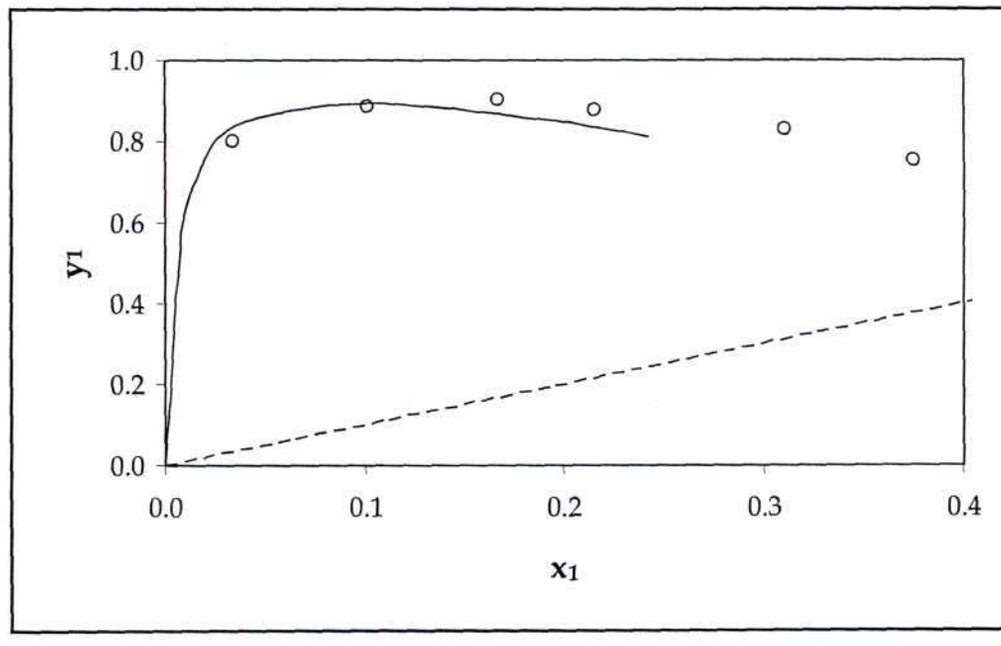

Figure 8.49: Plot of *x-y* data for methane (1) + sulfur dioxide (2) at 301.48 K. The solid line represents the experimental data and the symbols the simulation results.

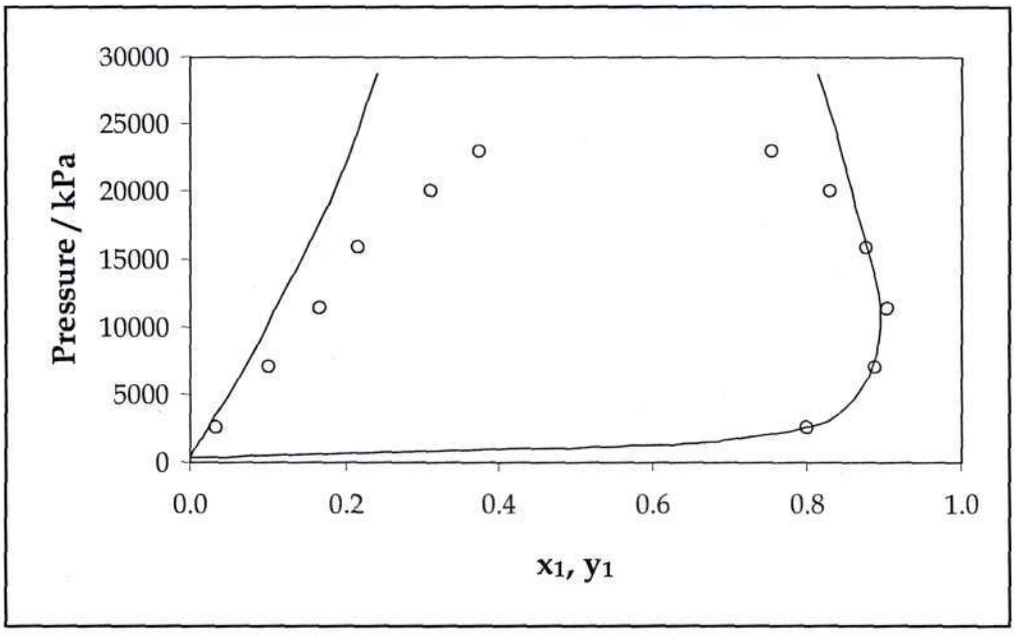

Figure 8.50: Plot of *P-x-y* data for methane (1) + sulfur dioxide (2) at 301.48 K. The solid line represents the experimental data and the symbols the simulation results.

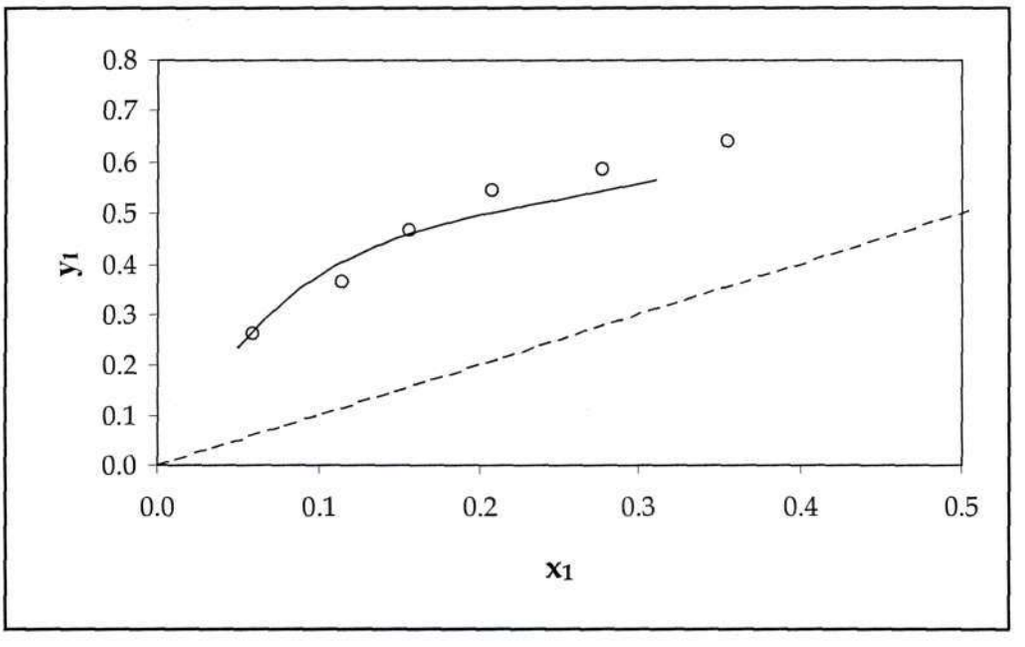

Figure 8.51: Plot of *x-y* data for propane (1) + sulfur dioxide (2) at 323.15 K. The solid line represents the experimental data and the symbols the simulation results.

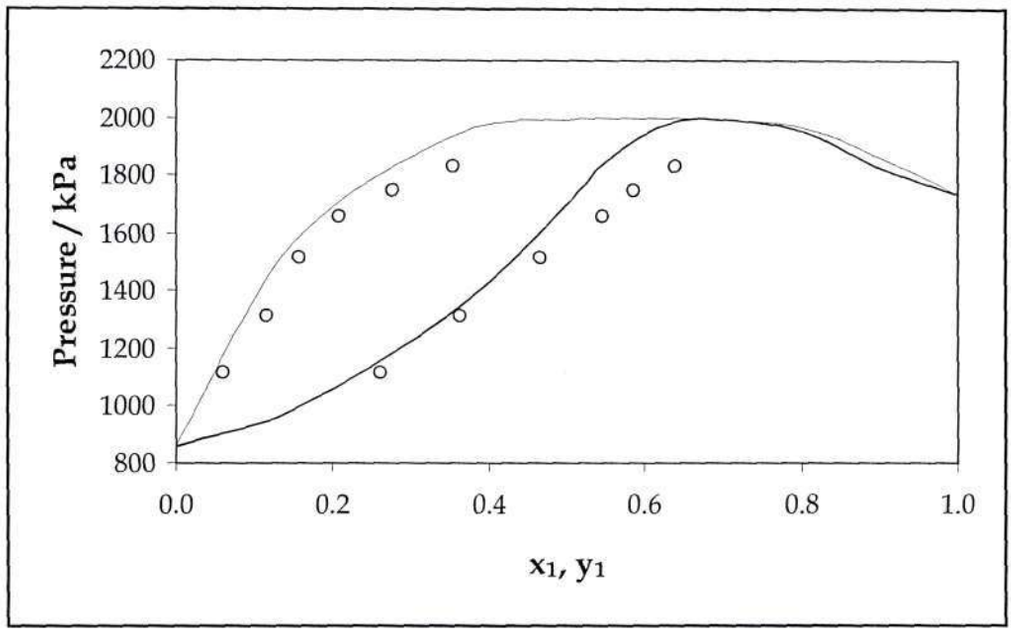

Figure 8.52: Plot of *P-x-y* data for propane (1) + sulfur dioxide (2) at 323.15 K. The solid line represents the experimental data and the symbols the simulation results.

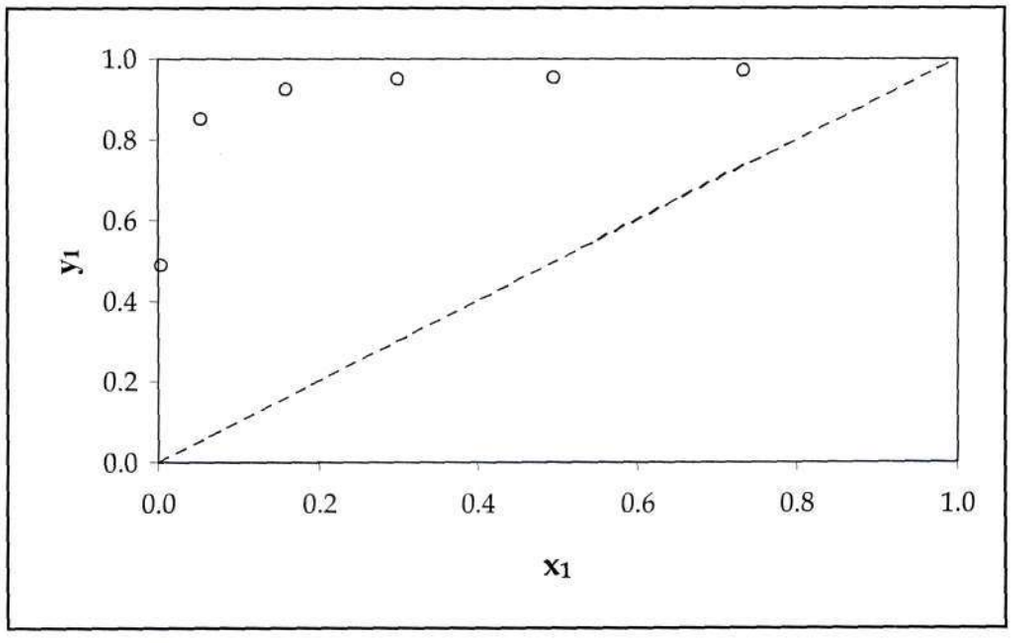

Figure 8.53: Plot of *x-y* data for sulfur dioxide (1) + ethanol (2) at 293.15 K.

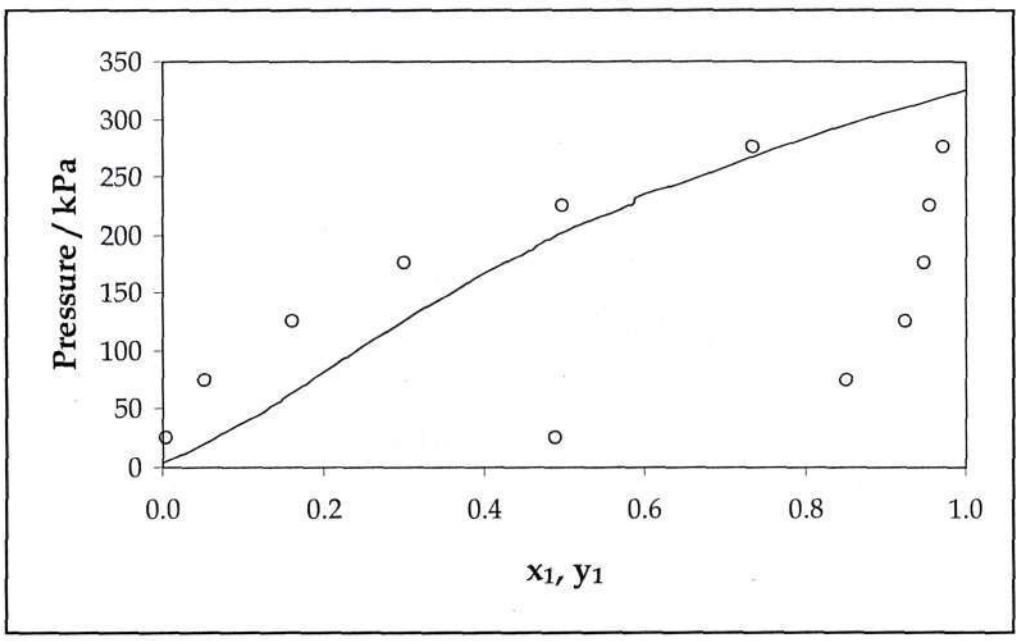

Figure 8.54: Plot of *P-x-y* data for sulfur dioxide (1) + ethanol (2) at 293.15 K. The solid line represents the experimental data and the symbols the simulation results.

Conversely, it appears from Figure 8.50 that the azeotropic composition is predicted with a fair degree of precision, but is slightly too high. This conclusion is reinforced by Figure 8.49 which clearly illustrates that the simulated compositions will cross the dashed 45° azeotrope-composition line at a higher point than the experimental data.

For the propane + SO_2 system at 323.15 K, the simulation results are plotted in Figures 8.51 and 8.52 with the measured data of Glòwka (1972) used for comparison. From the *x-y* and *P-x-y* plots it is clear that both the liquid and vapour compositions are over-predicted and that this becomes steadily worse as the pressure is increased. Hence, the experimental azeotropic composition (approximately 0.68) is also over-estimated. However, the system pressures are accurately predicted, within the statistical deviation of the simulation results. The simulated azeotropic pressure, therefore, should compare well with the measured value of 2000 kPa.

The final system involving SO_2 that was simulated was the SO_2 + ethanol system at 293.15 K (see Figures 8.43 and 8.54). The only experimental data that could be obtained for this system was that of Nitta et al. (1973, 1976) and unfortunately they only measured *P-x* data, making quantitative comparison with the simulation results difficult. What is clearly evident is that the system pressures are once-again over-predicted. As for the H_2S + methanol system discussed above, this over-estimation

lessens as the alcohol mole fraction decreases, indicating that this effect is largely a result of the alcohol partial pressures being over-predicted. As was observed for the H_2S + methanol binary, it is likely that the liquid compositions are being under-predicted, exaggerating the effect. However, the conclusions drawn remain ambiguous since an experimental x-y curve is not available for comparison in Figure 8.53.

It was found that for all of the binary VLE simulations involving SO_2, the Waldman and Hagler combining rule (Waldman and Hagler, 1993) provided the best prediction of the experimental data. The systems chosen for SO_2 differed from those for H_2S only because experimental data for SO_2 with ethane and methanol could not be obtained.

8.2.5 Sensitivity Analysis

The unlike interaction parameters are of considerable importance for any Lennard-Jones-based force field. Chen et al. (2001) attribute their over-prediction of azeotropic pressures for alkane + alcohol binary systems to their unlike-pair interactions being too weak. Despite the fact that the vital role that these parameters play in determining simulation outcomes is acknowledged in the literature, many researchers persist in using simple, empirical combining rules (e.g. Lorentz-Berthelot) to calculate unlike interaction parameters in mixtures. By demonstrating the considerable influence that these parameters can have on the simulation results, it is hoped that researchers will begin to take greater care in the selection of combining rules and the development of new, theoretically sound rules.

8.2.5.1 Lennard-Jones Parameters

A sensitivity analysis of the Lennard-Jones parameters was performed during the parameter optimization, where the response of the simulation results to small adjustments in these parameters was analyzed. It was found that even a slight perturbation of certain parameters (sometimes less than 1%) led to substantial deviations in the simulated pressures and densities. Although changes to all of the Lennard-Jones parameters produced similar fluctuations in the simulation results, the effect of the unlike interaction parameters will be focused on here.

Table 8.21 and Figures 8.41 to 8.44 show the discrepancy obtained for perturbations of approximately 5 % in the Lennard-Jones energy (ε_{NO}) and size (σ_{NO}) parameters of NO_2, respectively. It is clear that the effect of varying σ_{NO} has a far greater influence on the simulated results, but the deviations resulting from changes to ε_{NO} are also significant. A similar trend was observed for all three compounds parameterized in this project: the size parameters, σ, exerted a larger influence on the simulation output than the energy parameters, for a particular perturbation.

This outcome indicates that the selection of applicable combining rules is of considerable importance when simulating binary mixtures and may drastically alter simulation results. This effect is most pronounced when the Lennard-Jones parameters differ greatly (e.g. the parameters for H versus those for S for H_2S) and for all-atom type force fields where the sizes of atoms often differ considerably. Several authors such as Delhommelle and Millié (2001), Potoff et al. (1999) and Sadus (1993) have previously stressed the necessity of selecting appropriate combining rules. This sensitivity analysis emphasizes and quantifies the extent to which minor variations in the unlike interaction parameters can affect simulated data.

Table 8.21: Deviations in simulation results for perturbations of approximately 5 % in the Lennard-Jones energy (ε_{NO}) and size (σ_{NO}) parameters of NO_2[b]

σ_{NO} = 3.800 Å			σ_{NO} = 3.900 Å			σ_{NO} = 4.000 Å		
P	ρ_v	ρ_l	P	ρ_v	ρ_l	P	ρ_v	ρ_l
381.4	0.007	1.544	613.0	0.012	1.434	1091.7	0.022	1.324
875.8	0.016	1.483	1200.8	0.022	1.366	1975.4	0.039	1.242
1503.0	0.027	1.417	2291.7	0.042	1.300	3693.6	0.077	1.179
2866.0	0.047	1.351	3907.8	0.073	1.218	5530.2	0.113	1.045
4235.6	0.077	1.279	5732.7	0.114	1.115	8937.9	0.208	0.935
ε_{NO} = 1.750 K			ε_{NO} = 1.850 K			ε_{NO} = 1.950 K		
P	ρ_v	ρ_l	P	ρ_v	ρ_l	P	ρ_v	ρ_l
630.4	0.012	1.458	613.0	0.012	1.434	693.2	0.014	1.433
1312.5	0.024	1.394	1200.8	0.022	1.366	1438.0	0.027	1.362
2143.9	0.039	1.318	2291.7	0.042	1.300	2557.5	0.047	1.286
3782.7	0.073	1.25	3907.8	0.073	1.218	3760.8	0.076	1.202
5911.9	0.113	1.161	5732.7	0.114	1.115	6484.1	0.123	1.107

[b] Pressures and saturated densities are given in units of kPa and g/ml, respectively.

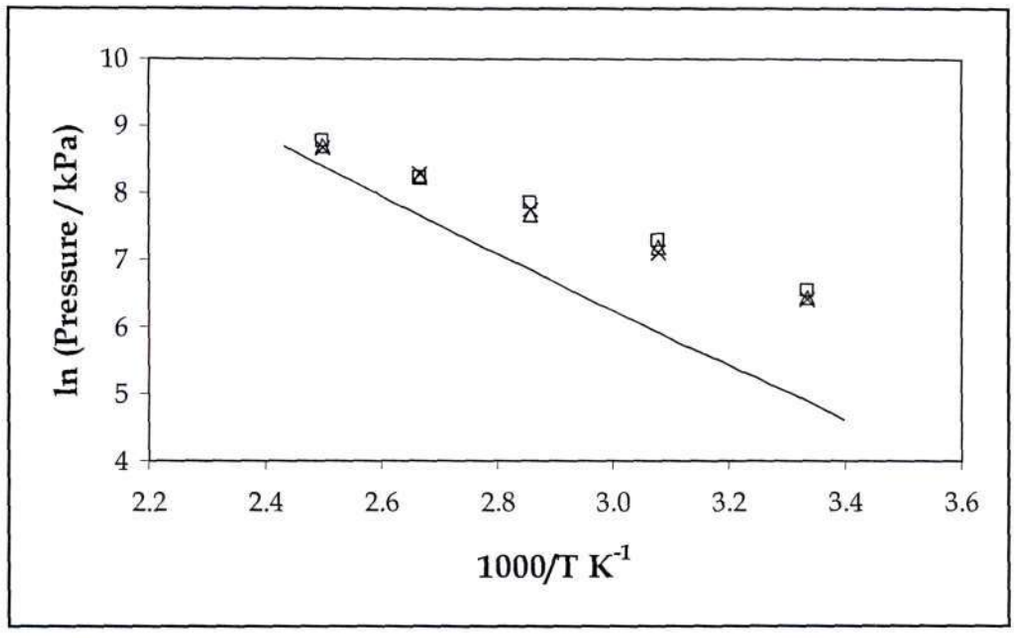

Figure 8.55: Deviation in the saturated vapour pressure simulation results following a perturbation of approximately 5 % in the Lennard-Jones energy parameter (ε_{NO}) of NO_2. Symbols represent simulation data for ε_{NO} values of 1.750 (Δ), 1.850 (x) and 1.950 (□), the line is experimental data.

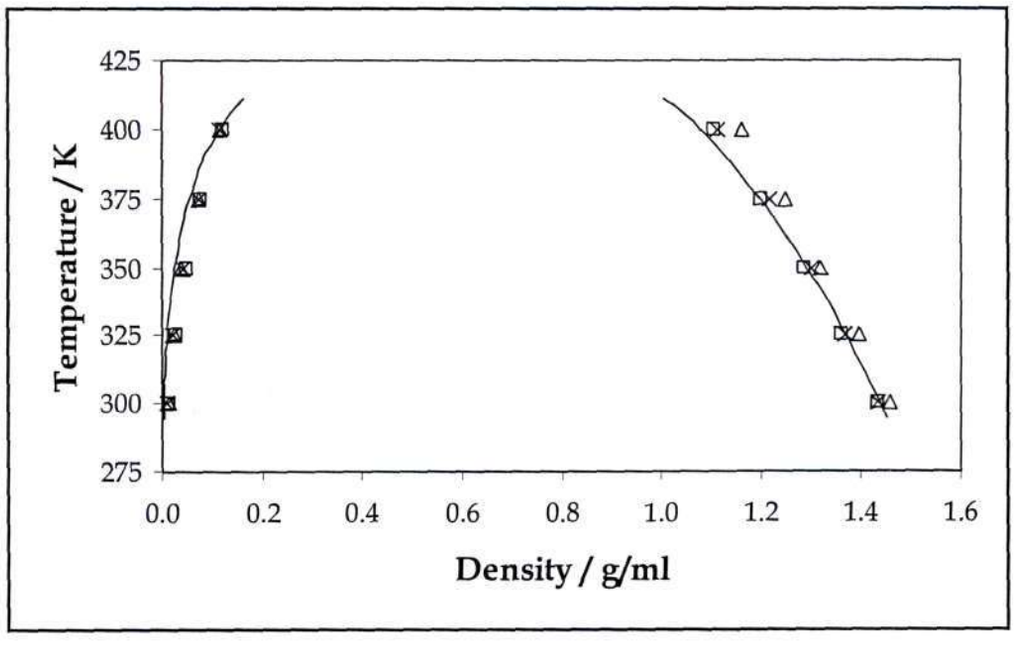

Figure 8.56: Deviation in the saturated density simulation results following a perturbation of approximately 5 % in the Lennard-Jones energy parameter (ε_{NO}) of NO_2. Symbols represent simulation data for ε_{NO} values of 1.750 (Δ), 1.850 (x) and 1.950 (□), the line is experimental data.

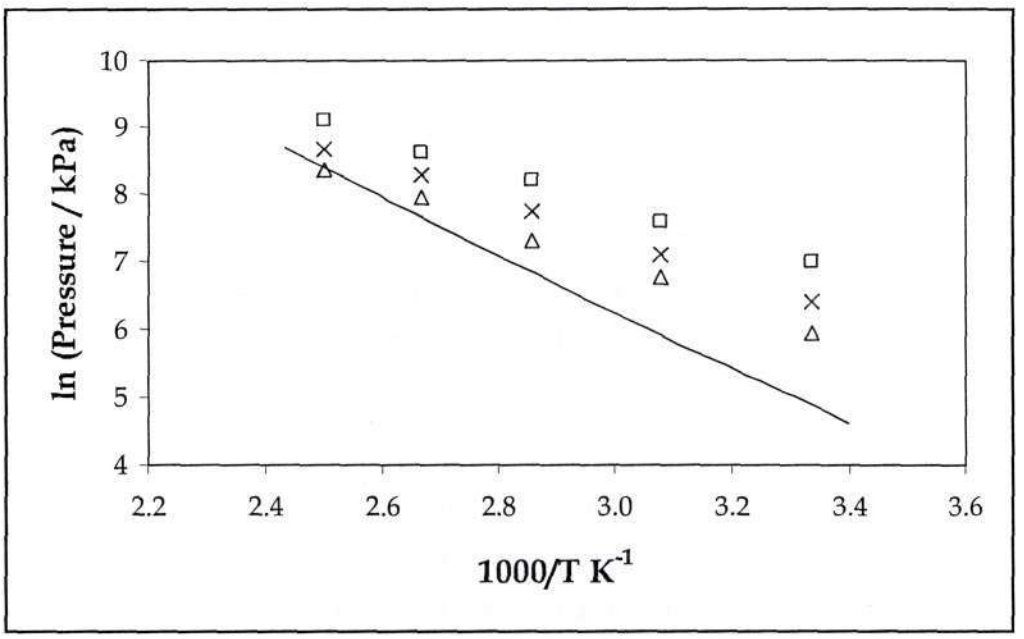

Figure 8.57: Deviation in the saturated vapour pressure simulation results following a perturbation of approximately 5 % in the Lennard-Jones size parameter (σ_{NO}) of NO_2. Symbols represent simulation data for σ_{NO} values of 3.800 (Δ), 3.900 (x) and 4.000 (□), the line is experimental data.

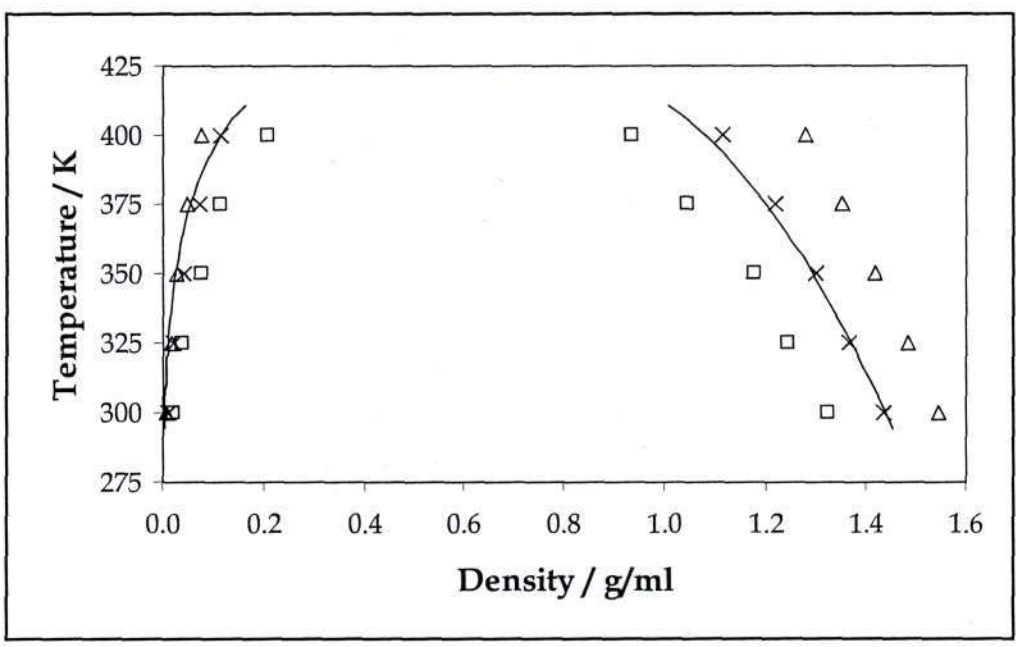

Figure 8.58: Deviation in the saturated density simulation results following a perturbation of approximately 5 % in the Lennard-Jones size parameter (σ_{NO}) of NO_2. Symbols represent simulation data for σ_{NO} values of 3.800 (Δ), 3.900 (x) and 4.000 (□), the line is experimental data.

8.2.5.2 *Combining Rules*

The unexpectedly sharp response of the simulation results to minor perturbations in the Lennard-Jones parameters led to an investigation into the combining rules used to calculate unlike interaction parameters when simulating mixtures. In this work, these combining rules were used to determine parameters for the unlike interactions between pseudo-atoms from the TraPPE-UA force field and the H_2S and SO_2 atoms involved in the binary phase equilibrium simulations discussed in Sections 8.2.4.2 and 8.2.4.3.

The methane + H_2S system was selected for analysis of the effect that the different combining rules have on the simulation results. Thereafter, the most applicable mixing rule for a particular binary system (determined by running initial simulations using each of the five different mixing rules at the start of each system) was used in each case. It was found that a particular mixing rule would provide the best prediction for a series of systems involving a particular compound: for H_2S systems, the Halgren combining rule (1992) proved best, and for SO_2 it was the combining rule of Waldman and Hagler (1993). The exception was the ethane + H_2S system, which is discussed further below.

Figure 8.59 shows the simulation results for the methane + H_2S system using the following combining rules: Lorentz-Berthelot (Lorentz, 1881; Berthelot, 1898), geometric, Kong (1973), Waldman-Hagler (1993) and Halgren (1992). Details regarding these mixing rules may be found in Section 5.2.3. It is clear from Figure 8.59 that the Halgren mixing rule provides the best prediction of the experimental data, followed by the Waldman-Hagler, Kong, geometric and Lorentz-Berthelot rules. The results obtained using the geometric and Lorentz-Berthelot rules are practically indistinguishable, although the vapour phase mole fractions are not quite as badly over-predicted by the geometric mixing rule. However, it is obvious that altering the mixing rule has a significant effect on the outcome of the simulations and careful selection of the appropriate combining rule for a particular system is of considerable importance.

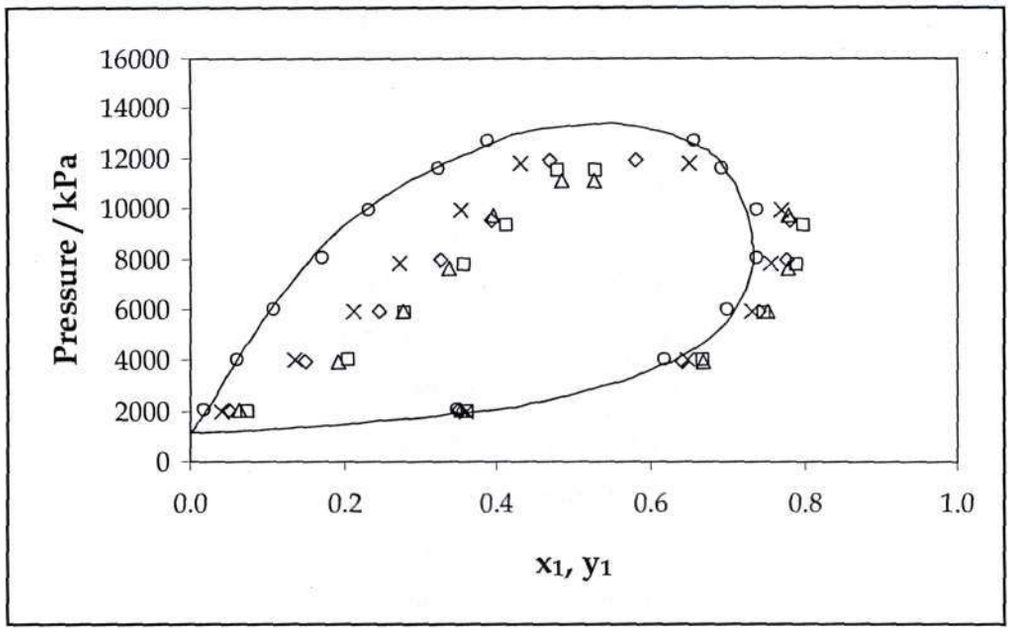

Figure 8.59: Simulation results for the methane (1) + hydrogen sulfide (2) system at 277.59 K. The solid line represents experimental data and the symbols are the simulation data obtained using the different combining rules: Lorentz-Berthelot (□), geometric (△), Kong (◊), Waldman-Hagler (x) and Halgren (○).

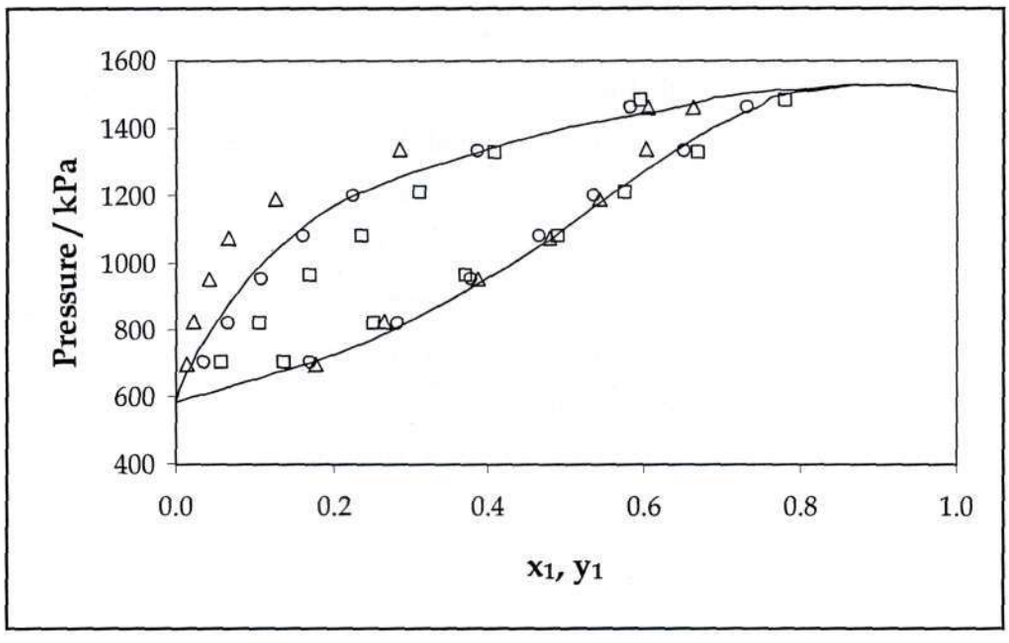

Figure 8.60: Simulation results for the ethane (1) + hydrogen sulfide (2) system at 255.32 K. The solid line represents experimental data and the symbols are the simulation data obtained.

It is noticeable that the popular and commonly-used Lorentz-Berthelot mixing rule fairs poorest in this instance. Considering that the TraPPE-UA force field (used for the alkanes and alcohols involved in the binary VLE simulations) was parameterized using this mixing rule, this result is especially disquieting.

It is interesting to examine the extent to which the unlike interaction parameters (calculated using the various combining rules) differ from one another for the methane + H_2S system. To this end, the Lorentz-Berthelot parameters are used as a yardstick to which the other mixing rule parameters are compared. The S–CH_4 size and energy parameters deviate by less than 1 and 5 %, respectively. Although these fluctuations appear rather small for S, even a variation of 5 % in the Lennard-Jones parameters could have a significant effect on the simulation results, as was demonstrated by the sensitivity analysis conducted during the parameterization of H_2S, SO_2 and NO_2. For the H–CH_4 parameters, the size and energy parameters deviate by as much as 50 and 100 %, respectively. This emphasizes that force fields containing atoms like hydrogen (with relatively small Lennard-Jones parameters) will always be susceptible to large variations in the unlike interaction parameters depending on the combining rule used.

Figure 8.60 shows the simulation results for the ethane + H_2S system. The triangles and squares represent the results obtained from the Halgren and Waldman-Hagler combining rules, respectively. The actual simulation results presented in Figure 8.44 (Section 8.2.4.2) are again shown here as circles and were obtained using a combined mixing rule. It was observed that the Halgren mixing rule was under-predicting the mole fractions (particularly the liquid phase) and, in contrast, the Waldman-Halgren mixing rule was over-predicting the mole fractions. Hence, it was decided to investigate the effect of using the Lorentz-Berthelot mixing rule to combine the unlike interaction parameters obtained from the other two rules. The simulation results from this combined mixing rule compare very well with the experimental data. Interestingly, the effect on the liquid phase mole fraction is pronounced, whereas the vapour phase mole fractions remain almost indistinguishable until the mole fractions are greater than 0.5 and the ethane concentration begins to dominate.

The results presented here in Section 8.2.5 unambiguously demonstrate that the simulation results are very sensitive to the combining laws. If this difficulty is to be overcome, a combining rule that *a priori* provides a satisfactory prediction for any set of mixtures (where the model for each component that exists in the mixture has been

developed independently) is essential. However, it is by no means certain that the development of such a combining rule is mathematically possible. In the meantime, the only viable approach is to use a suitable combining rule when fitting a potential model to experimental data, and then consistently use this rule with the optimized model.

8.2.6 Association Effects

The polar compounds studied during this project all exhibit certain association effects as a result of the charged nature of their atoms. These effects include phenomena such as hydrogen bonding and dimerization, which are of considerable interest to researchers because of the influence they exert on the chemical and physical behaviour of these compounds. Molecular simulation offers a unique facility for studying the microscopic structures of polar molecules through the analysis of intermolecular radial distribution functions. In the subsections that follow, the radial distributions are all taken from simulations run at a reduced temperature of approximately 80 %.

8.2.6.1 Alcohols

The site to site radial distributions for 2-methyl propanol (Figure 8.61) and 1-pentanol (Figure 8.62) clearly show that hydrogen bonding is occurring. Figures 8.61 and 8.62 compare well with the radial distribution functions of Chen et al. (2001), with the peaks occurring between 2.7 and 2.8 Å. This agrees well with experiments using X-ray diffraction to determine the microstructure of alcohol systems (Chen et al., 2001).

It is evident from the figures presented below that the response (or peak) increases in magnitude as the molecular size increases. The same trend is shown by the radial distribution functions of Chen et al. (2001). The truncation of the potential at 14 Å is justified by Figures 8.61 and 8.62, which show that the alcohol radial distributions have settled down to a value of one, well before this cut-off radius is reached.

Figure 8.63 shows a snapshot of the 1-propanol system at 400 K produced using the RasMol visualization package (Bernstein, 2001). It is obvious that there is aggregation occurring due to hydrogen bonding between the alcohol molecules.

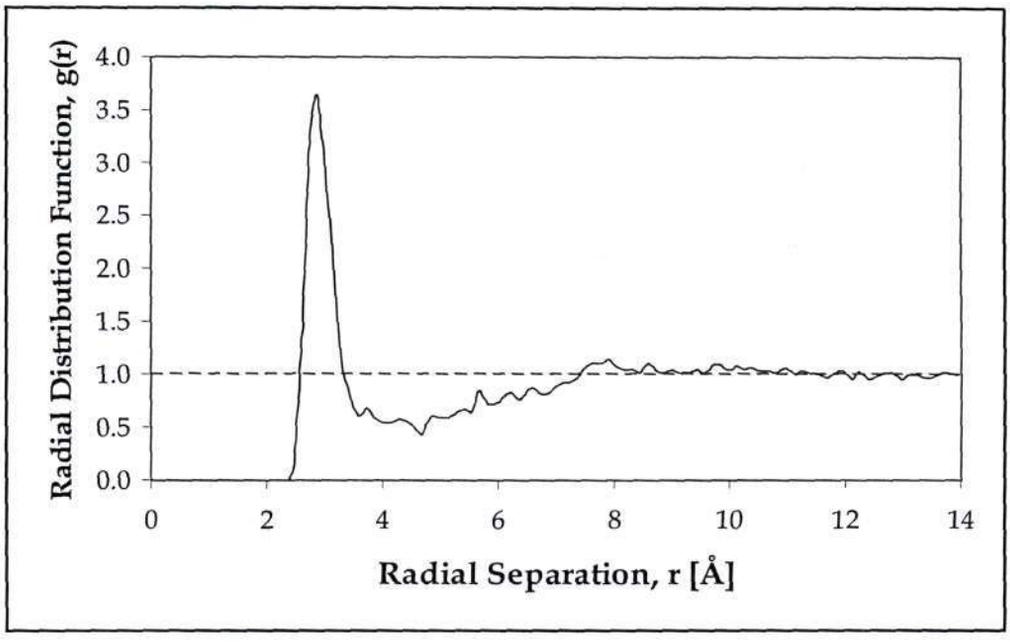

Figure 8.61: Radial distribution function of oxygen - oxygen in 2-methyl-1-propanol.

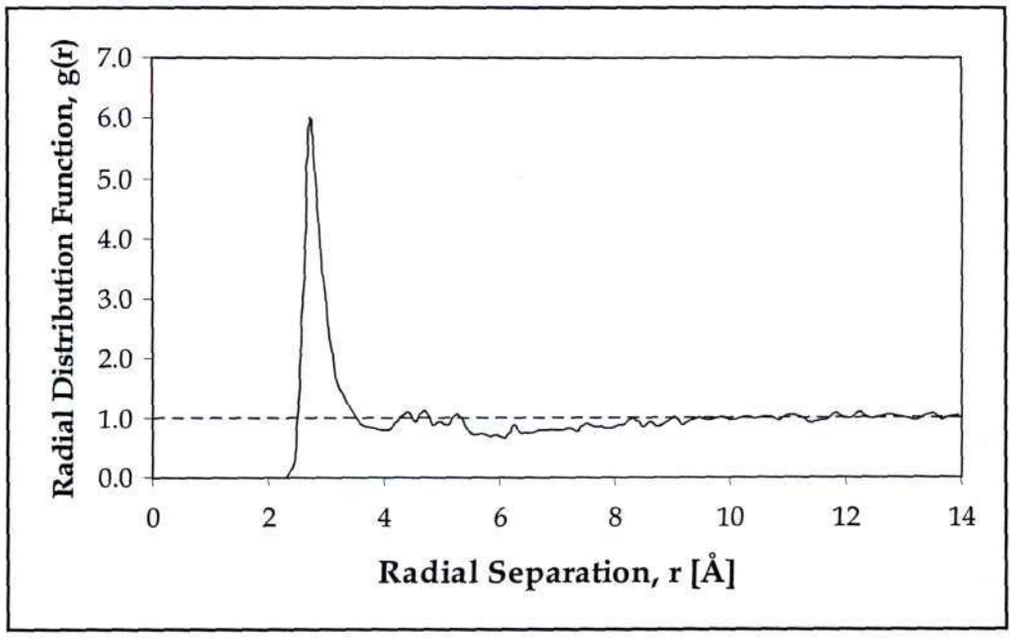

Figure 8.62: Radial distribution function of oxygen - oxygen in 1-pentanol.

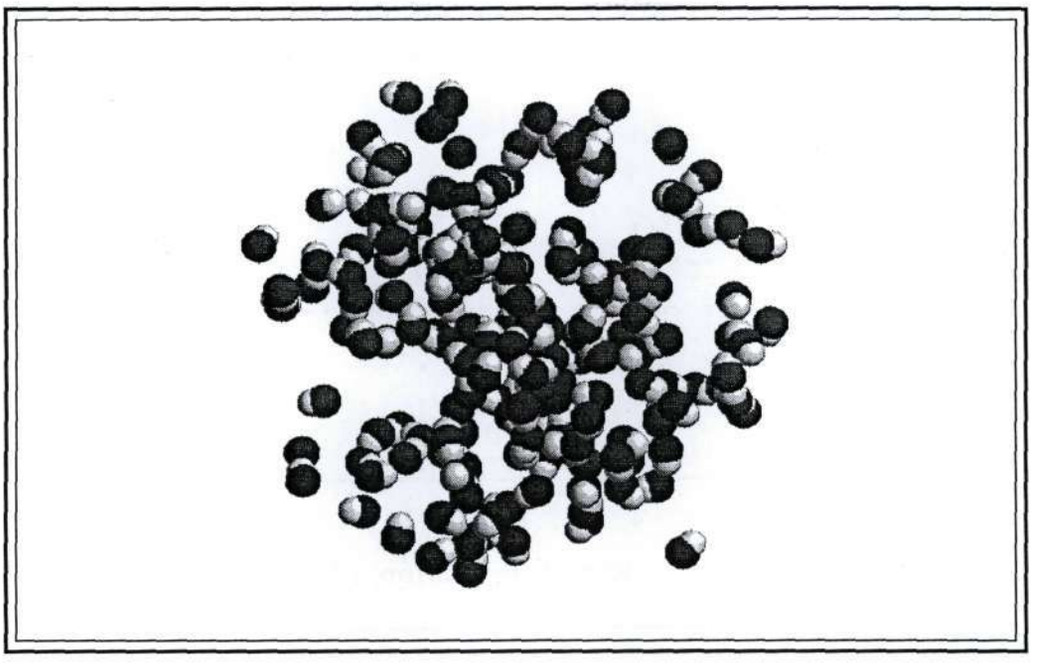

Figure 8.63: Snapshot of the liquid phase of the 1-propanol system at 400 K. The alkyl tails have been removed to improve clarity.

8.2.6.2 *Carboxylic Acids*

The intermolecular radial distribution functions of the carboxyl oxygen - hydroxyl hydrogen for acetic acid and pentanoic acid are shown in Figures 8.64 and 8.65, respectively. Both plots show a strong peak that occurs at a separation of approximately 2 Å, indicating the existence of dimers formed by strong hydrogen bonds. Figure 8.66 shows a snapshot of the simulated vapour phase of acetic acid, with a number of dimers visible. The radial distribution function for acetic acid closely resembles the plots of Kamath et al. (2004), the only noticeable difference being the size of the peak. This minor disparity is attributed to the differing simulation conditions.

As was found for the alcohols, the radial distribution function peak is larger for the larger acid. This is not an indication that dimerization occurs more readily for the bulkier acid. In fact, the opposite trend should exist, with the steric hindrance of the larger alkyl tail making dimer formation more difficult as the acid size increases.

Again, both radial distribution functions quite obviously reach an effective value of unity well before the cut-off radius is reached at 14 Å.

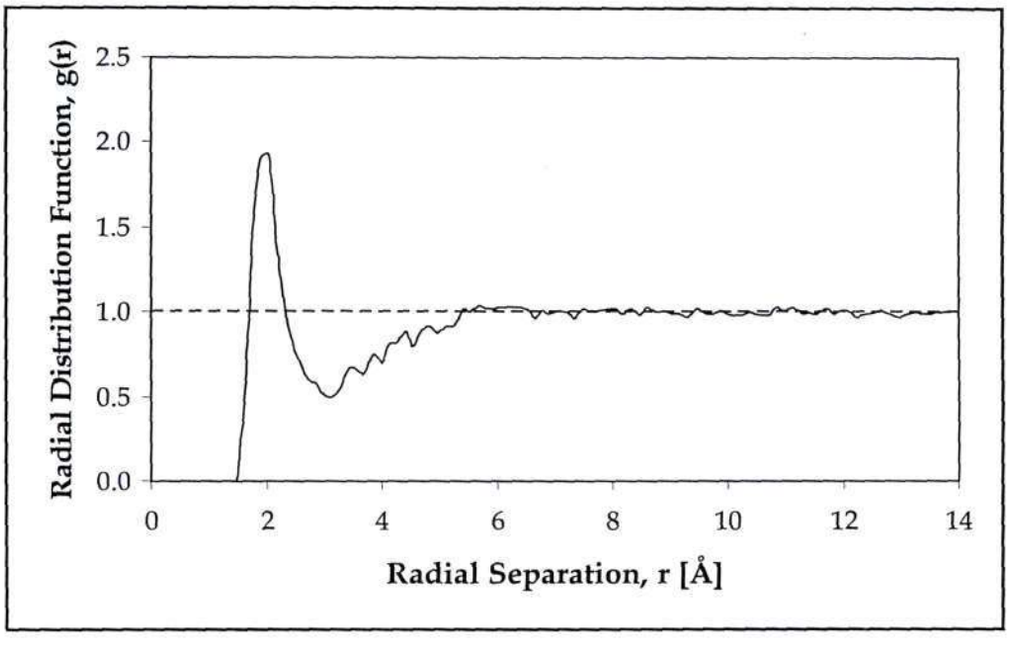

Figure 8.64: Radial distribution function of carboxyl oxygen - hydroxyl hydrogen in acetic acid.

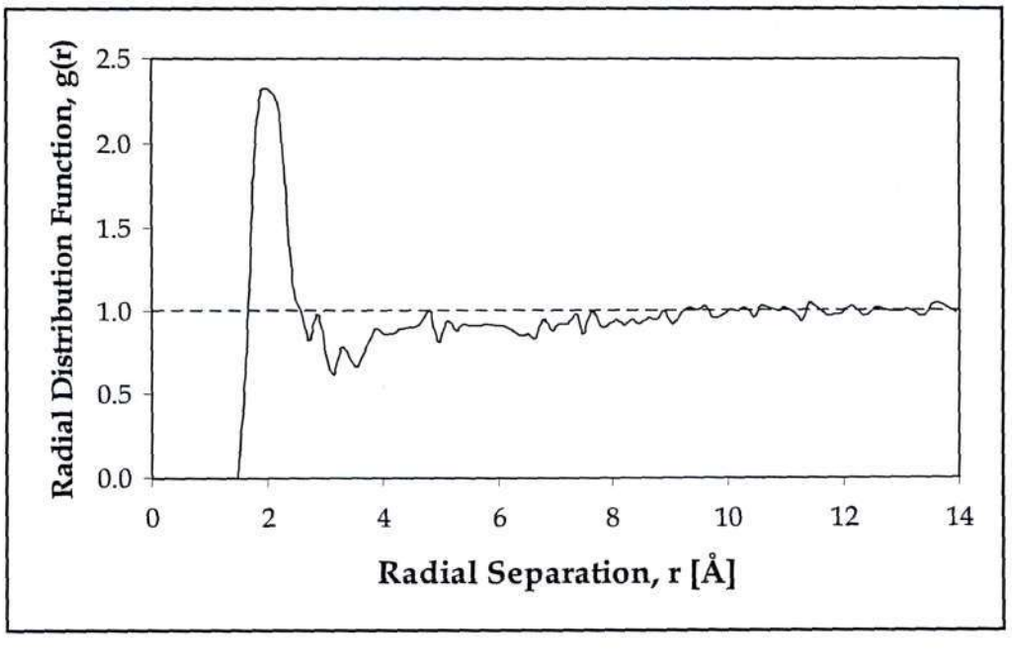

Figure 8.65: Radial distribution function of carboxyl oxygen - hydroxyl hydrogen in pentanoic acid.

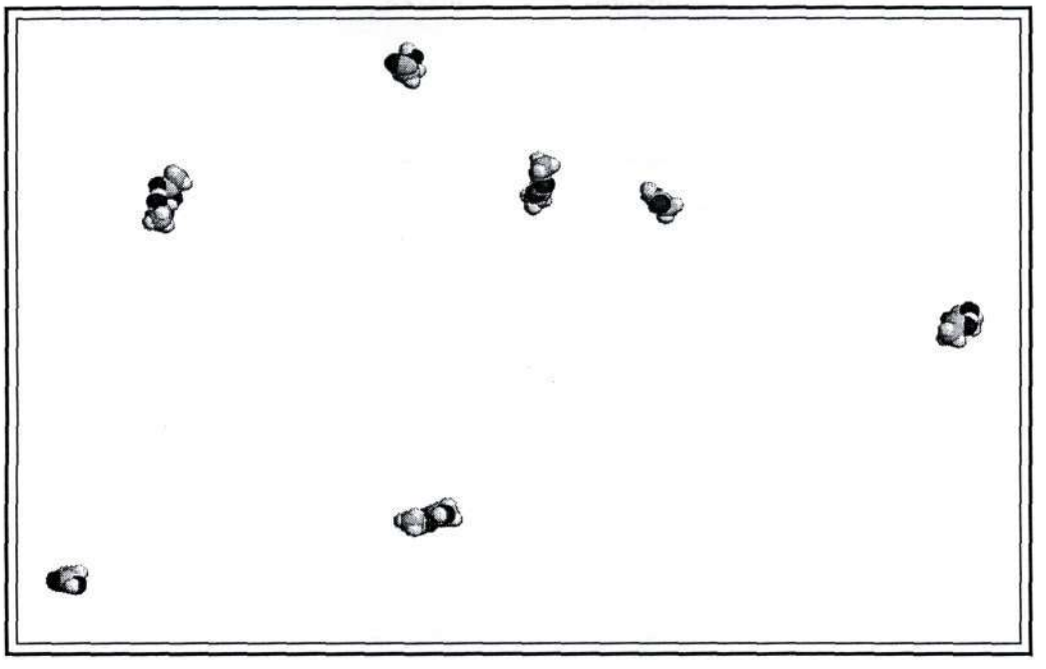

Figure 8.66: Snapshot of the acetic acid vapour phase at 300 K, showing the extensive dimerization that occurs in carboxylic acid systems.

8.2.6.3 *H_2S, SO_2 and NO_2*

The close proximity of charged atoms characteristic of the formation of dimers and hydrogen-bonded complexes is not evident in the H_2S, SO_2 and NO_2 radial distribution functions (Figures 8.67 to 8.69), indicating that these association effects are not occurring or are occurring very weakly in these simulations. This result holds particular significance for NO_2 since we know from experimental results that this molecule dimerizes extensively in both phases (as was discussed previously in Section 8.2.3.2). The lack of dimerization behaviour is one of the major reasons why the NO_2 parameterization simulation results deviated from experimental data to such an extent. Figure 8.69 demonstrates that no dimers are being predicted and that modifications to the model (and/or the partial charges) used here are necessary if these results are to be improved upon (see Chapter 10).

The plateau that occurs in the SO_2 radial distribution is very interesting. A similar anomaly was observed by Kamath at al. (2004) and indicated a small chance of some associating-type interactions occurring. This is attributed to the fact that the charges on

the SO_2 molecule are significantly larger than those of H_2S or NO_2, allowing for the possibility of some association between the SO_2 molecules.

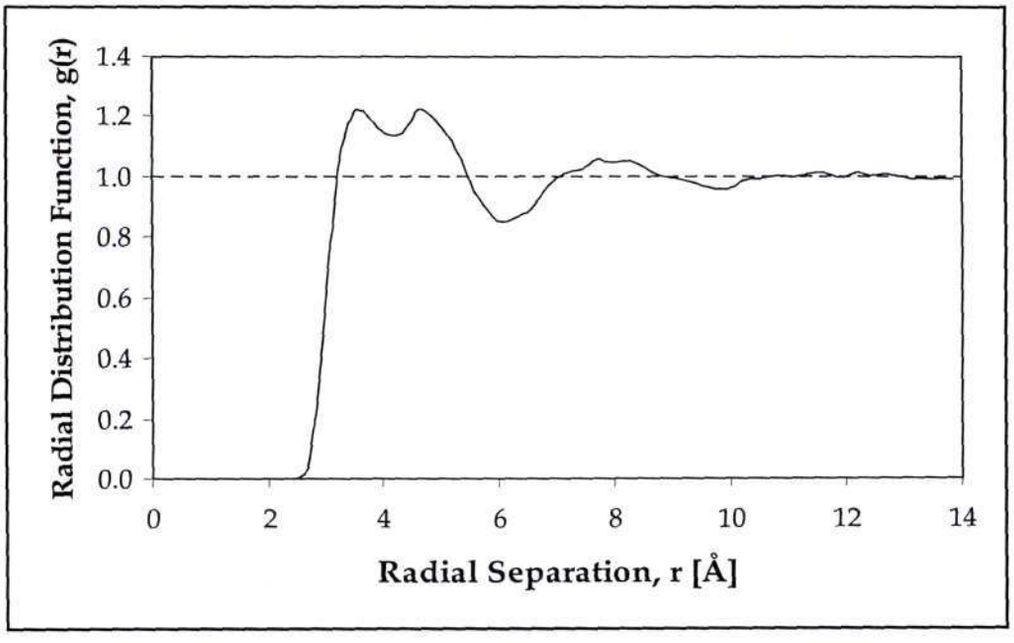

Figure 8.67: Radial distribution function of hydrogen - sulfur in hydrogen sulfide.

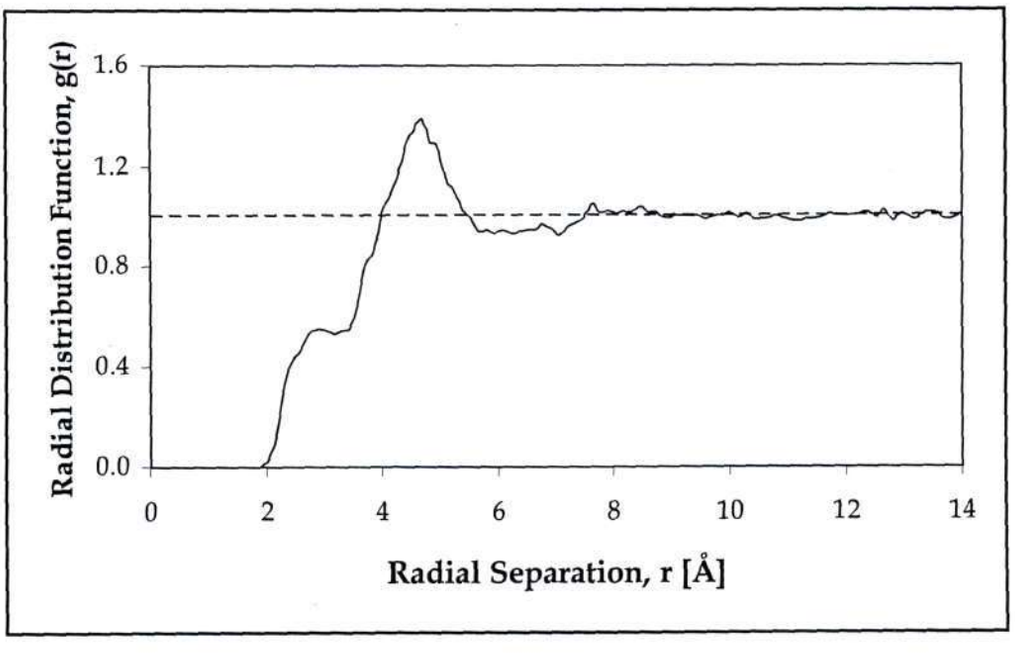

Figure 8.68: Radial distribution function of oxygen - sulfur in sulfur dioxide.

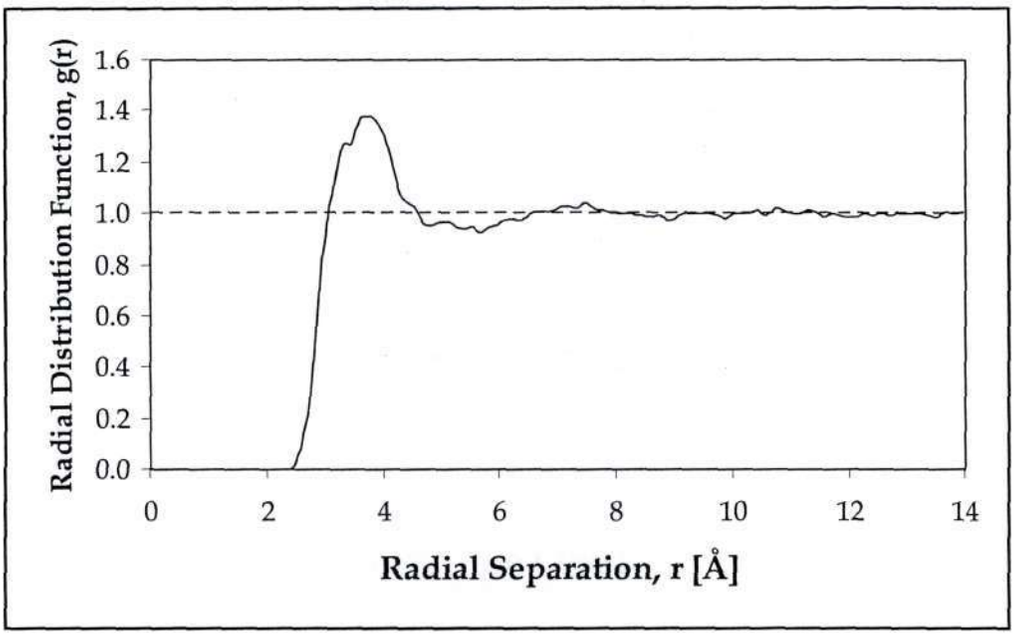

Figure 8.69: Radial distribution function of oxygen - nitrogen in nitrogen dioxide.

8.2.6.4 Propanoic Acid + Pentanoic Acid Binary VLE System

One of the central issues arising from the chemical theory modeling portion of the project was how to calculate the heterodimerization constant, K_{AB}. Two possibilities existed, one of which was a geometric mean of the pure component dimerization constants K_{AA} and K_{BB} (Equation 2.21), and the other was twice the geometric mean (Equation 2.20), implying that the heterodimer was twice as likely to form as either of the homodimers, provided the dimerization constants are of a similar magnitude. Through empirical investigation undertaken during the modeling phase of this study, it was found that Equation 2.20 resulted in physically meaningless activity coefficients, whilst Equation 2.21 produced good results. The proficiency of molecular simulation in analyzing the microstructure of chemical systems provided an ideal opportunity to test this empirical result.

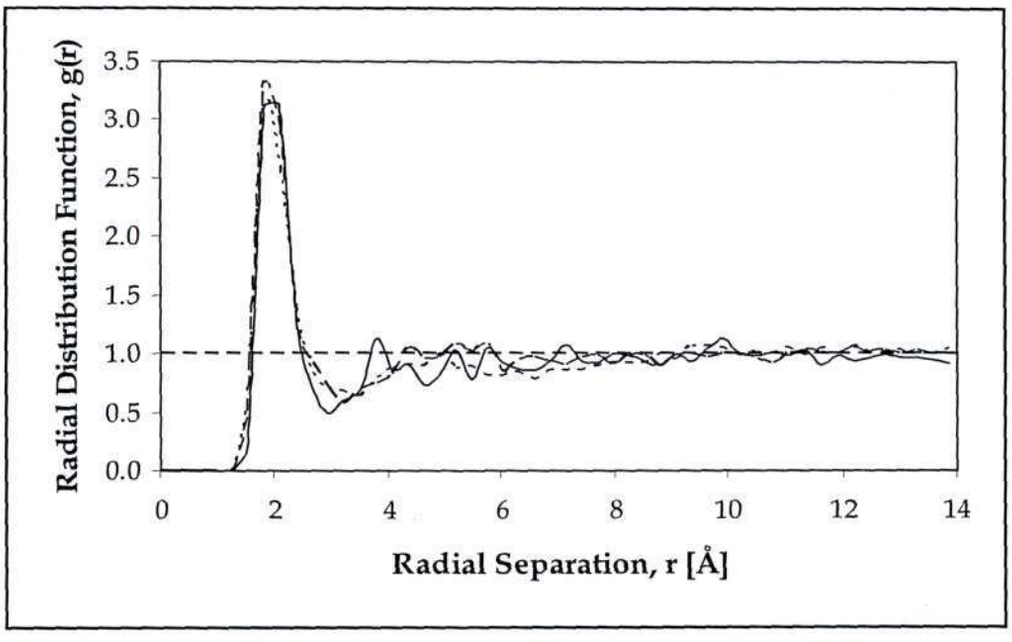

Figure 8.70: Radial distribution functions for the propanoic acid + pentanoic acid system at 393 K. The solid line represents the propanoic acid carboxyl oxygen - propanoic acid hydroxyl hydrogen (i.e. propanoic acid homodimer), the dashed line represents the propanoic acid carboxyl oxygen - pentanoic acid hydroxyl hydrogen (i.e. heterodimer) and the dotted line represents the pentanoic acid carboxyl oxygen - pentanoic acid hydroxyl hydrogen (i.e. pentanoic acid homodimer).

The propanoic acid + pentanoic acid system at 393 K was used to study the dimerization phenomenon. A simulation that contained a roughly equal number of propanoic acid and pentanoic acid molecules was chosen so that the results would not be biased. The intermolecular radial distribution function (shown in Figure 8.70) was then analyzed and it is clear that the peaks for each type of possible dimer (i.e. the two homodimers and the heterodimer) are almost identical. This result indicates that the proportion of dimers is roughly equal and that no particular dimer is favoured to form preferentially. Hence, the simulation results support the use of Equation 2.21 rather than Equation 2.20.

9

Conclusion

The initial area of interest in this project was an inquiry into the modeling of carboxylic acid vapour-liquid equilibrium (VLE) data using chemical theory. This investigation followed on from the preceding Masters project involving laboratory measurement of carboxylic acid phase equilibrium data. It was found that the usual gamma-phi formulation of VLE was not adequate for the binary carboxylic acid systems and, hence, the chemical theory of vapour-phase imperfections was adopted to take the acid dimerization occurring in the vapour phase into account.

9.1 Chemical Theory Modeling

The carboxylic acid systems studied were propanoic acid + pentanoic acid and 2-methyl propanoic acid + pentanoic acid, with three isotherms at 393.15, 403.15 and 413.15 K, respectively, and an isobar at 20 kPa being modeled. In each case, liquid-phase activity coefficients were calculated using the chemical theory for the vapour phase, based largely on procedures developed by Tamir and Wisniak (1975). It was found that the best results were obtained by using a geometric mean average for the heterodimerization constant, K_{AB}. The systems all showed positive deviations from Raoult's law, although small negative deviations were observed in the dilute regions

and seem to be typical of binary systems containing only carboxylic acids (possibly due to substantial heterodimer formation).

The area test was used to determine the thermodynamic consistency of the measured data. Considering that the calculated area residuals reflect unavoidable errors from a number of sources (and the relatively ideal shape of the phase envelopes typical of binary systems involving like compounds), the residuals for the 2-methyl propanoic acid + pentanoic acid systems are remarkably low. The residuals for the propanoic acid systems are, in general, significantly higher and this is attributed to the formation of small amounts of propanoic acid trimer or tetramer not taken into account in the chemical theory.

To allow VLE data regression of carboxylic acid systems (that necessitate use of the chemical theory of the vapour phase), an innovative bubble-point regression algorithm was developed that incorporates chemical theory into the calculation procedure. This technique permits accurate determination of the system pressure and vapour phase composition throughout the liquid composition range at a particular temperature, allowing for design of accurate distillation equipment. The fit of the model to the measured data is excellent for the 2-methyl propanoic acid + pentanoic acid system, and completely acceptable for the propanoic acid + pentanoic acid system. The significant deviation from unity of the vapour-phase correction factors calculated during the fitting process confirms that the acid systems deviate considerably from ideal gas behaviour, despite the ideal appearance of the phase envelopes.

9.2 Molecular Simulation Results

The principle aim of this study was to examine the ability of currently available molecular simulation techniques to accurately predict phase equilibrium data of polar compounds. To this end, pure component data has been simulated for alcohols, carboxylic acids and (as part of the parameterization section of the project) hydrogen sulfide (H_2S), sulfur dioxide (SO_2) and nitrogen dioxide (NO_2). Critical properties for these compounds were then estimated by constrained fitting to the simulation data. Thereafter, binary VLE simulations were performed involving (principally) carboxylic acids, H_2S and SO_2. During the parameterization process and the binary simulations, a sensitivity analysis was conducted to determine the effect on the simulation results of

varying the combining rules usually used in calculating the unlike interaction parameters. Finally, radial distribution functions were used to investigate certain effects of association (such as hydrogen bonding and dimerization) that often occur in systems containing polar molecules.

9.2.1 Pure Component Simulation Results

Pure component simulations were performed for several alcohols and carboxylic acids, viz. methanol, 1-propanol, 2-methyl-1-propanol, 2-methyl-2-propanol, 1-pentanol, acetic acid, propanoic acid, 2-methyl propanoic acid and pentanoic acid. All the pure component systems studied showed the saturated vapour pressure over-prediction characteristic of the TraPPE-UA force field (varying between 5 and 20 % depending on the species). The saturated liquid density predictions were generally good, differing from experiment (on average) by approximately 1-2 %, whilst the saturated vapour densities were consistently over-predicted to some extent. It was observed during the parameterization of H_2S, SO_2 and NO_2 that the vapour pressure and density are closely dependent and, hence, if one property is over-estimated, the other will follow suit.

In general, the estimated critical properties compared well with experiment. For the alcohols, the critical temperature and density differed from experiment by 1.5 and 1.7 %, respectively. The values for the carboxylic acids were even better, with a deviation of less than 1 % on average. Both classes of compound displayed the anticipated pressure over-estimation, with their critical pressures considerably larger than experiment on average (16 and 21 % for the alcohols and acids, respectively). The Clausius-Clapeyron plots allowed for interpolation of the normal boiling points. The alcohol predictions were very good, deviating from experiment by less than 0.5 %; however, the carboxylic acid normal boiling points were under-estimated by as much as 4.7 %.

In order to calculate the critical properties, the simulated vapour-liquid coexistence data were fitted to the density scaling law and law of rectilinear diameters. During this process, the scaling exponent was allowed to vary slightly from the universal Ising value to obtain the best possible fit to the coexistence curve, critical temperature and critical density. It was found that a value of 0.29 for the alcohol and carboxylic acid

scaling exponents provided the best fit (except for methanol, where 0.28 gave better results).

Before the 2-methyl propanoic acid could be simulated, several parameters needed to be added to the TraPPE-UA force field since the force field had not previously been used to simulate branched carboxylic acids. The required bond angle and torsional parameters involving the -CH- functional group were borrowed from existing bonds and torsions within TraPPE-UA, extending the force field to include branched acids.

9.2.2 Parameterization of H_2S, SO_2 and NO_2

Parameters for a combined Lennard-Jones 12-6 plus Coulombic potential energy surface were developed for the polar molecules H_2S, SO_2 and NO_2 by fitting to experimental saturation densities. The simulation results for H_2S demonstrated that the proposed potential model and parameters delivered predictions of the H_2S phase behaviour of comparable accuracy to the latest force fields available in the literature.

SO_2 and NO_2 had not previously been parameterized for a Lennard-Jones 12-6 based force field. For SO_2, there was excellent agreement between the simulation results and experimental data. In the case of NO_2, the saturated liquid density predictions were very good, but the vapour pressures and densities were significantly over-predicted. This is likely to be a result of the potential model not being attractive enough. This was discussed in more detail in Chapter 8 where the lack of dimerization observed in the NO_2 simulations was also attributed to the atomic charges being too low. Various means of improving or altering the current potential model to take this into account have been put forward.

The critical properties were also estimated for H_2S, SO_2 and NO_2. For H_2S and SO_2, the estimated properties show good agreement with their experimental counterparts. For NO_2, the estimated critical temperature is accurate, but the density is, not surprisingly, over-predicted. Again, the scaling exponent was allowed to vary during the fitting procedure and the best fits were obtained for effective scaling exponents of 0.315, 0.325 and 0.290 for H_2S, SO_2 and NO_2, respectively. Interestingly, the value of β_{eff} for SO_2 is identical to the universal Ising value (0.325), indicating that the SO_2 coexistence curve is similar in shape to that of non-polar rare gases and alkanes.

9.2.3 Binary Phase Equilibrium Results

Binary component simulations involving carboxylic acids have not previously been attempted. In this work, the following systems were simulated: propanoic acid + pentanoic acid and 2-methyl propanoic acid + pentanoic acid, with an isobar at 20 kPa and isotherm at 393.15 K for each system. The compositions were accurately predicted, however, the vapour pressure over-estimation characteristic of the TraPPE-UA force field (mentioned previously) resulted in the *P-x-y* and *T-x-y* curves being over and under-predicted, respectively. These deviations are within the expected range of 15 - 40 kPa and 20 - 30 K (for the isobars and isotherms respectively), based on the departures from experimental data observed during the pure component carboxylic acid simulations.

The binary VLE systems for H_2S with methane, ethane, propane and methanol were, in general, remarkably accurate. The exception was the H_2S + methanol system in which the system pressure is over-estimated slightly. This occurred in conjunction with an under-prediction of the H_2S solubility in methanol, which exacerbated the effect. The azeotropes that occurred in the H_2S + alkane systems were predicted with a fair degree of accuracy.

Binary simulations for SO_2 with methane, propane and ethanol were also performed. The SO_2 + methane system displayed a systematic over-prediction of the liquid-phase compositions. Furthermore, it appears that the force fields used here (parameterized at low pressures) cannot handle the extreme system pressures resulting from the fact that methane is supercritical at this temperature (301.48 K) and all simulations attempted above a pressure of 25 MPa failed to separate into two distinguishable phases. Consequently, the simulated azeotropic pressure is far too low. For the propane system, both the liquid and the vapour compositions are over-predicted, with the effect worsening as the pressure increased. Consequently, the azeotropic composition is also higher than that measured experimentally. However, the system pressures were predicted relatively well and, naturally, so was the azeotropic pressure. For the SO_2 + ethanol system the only experimental data available was *P-x* data, making quantitative comparison with the simulation results difficult. However, the system appears to follow a similar trend to the H_2S + methanol system with the pressures over-predicted somewhat. Again, this lessens as the SO_2 concentration increases indicating that the

effect is caused by over-estimation of the alcohol partial pressures by the TraPPE-UA force field.

9.2.4 Sensitivity Analysis

The sensitivity of the simulation results to variation of the unlike interaction parameters (commonly determined through the use of simple combining rules) was studied during this project. The parameterization work permitted investigation of the effect of slight perturbations (5 %) to the actual Lennard-Jones parameters and the use of different combining rules (when simulating the methane + H_2S binary system) illustrated the impact of these mixing rules on the final simulation output. It was found that even slight changes to the unlike parameters can have a tremendous influence on the simulation results. This emphasizes the need for researchers to carefully consider the combining rule that they choose to employ. Furthermore, this result highlights the need for the development of new, theoretically sound mixing rules, as opposed to the popular, empirical rules that are frequently used.

9.2.5 Association Effects

The analysis of radial distribution functions may be used to identify association-type interactions occurring in molecular simulations. The alcohol and carboxylic acid radial distributions showed the anticipated strong peaks at relatively short molecular separations that reveal hydrogen bonding and dimerization, respectively. The results for H_2S, SO_2 and NO_2 indicated that almost no association was taking place, aside from a slight plateau in the SO_2 radial distribution function. This effect was attributed to the atomic charges of SO_2 being significantly larger than those of H_2S and NO_2. The fact that NO_2 showed no obvious signs of any intermolecular association interactions confirmed that the simulations were not reproducing the dimerization known to occur from experiment.

Analysis of the radial distribution function generated from data collected during one of the simulations performed for the propanoic acid + pentanoic acid system at 393 K proved that the distribution of dimers in a binary carboxylic acid system was divided

more or less equally between the two homodimers and the heterodimer. This result supports the decision to use the geometric mean (rather than twice the geometric mean) of the pure acid dimerization constants to calculate the heterodimerization constant during the chemical theory modeling.

9.3 Project Significance

The initial phase of the project demonstrated that in order to adequately describe a binary carboxylic acid VLE system, the chemical theory of vapour phase imperfections must be applied. It was found that the conventional gamma-phi formulation of VLE did not satisfactorily account for the dimerization occurring extensively in these acid systems. Application of the chemical theory permitted accurate calculation of the experimental liquid-phase activity coefficients, which could then be fitted to a Gibbs excess energy model such as the TK-Wilson equation and regressed using a novel bubble-point calculation scheme incorporating chemical theory.

The principle project focus was the Monte Carlo simulation of polar compounds. Charged and polar species are frequently encountered in industry, and accurate phase equilibrium data is vital for the efficient running of plant processes involving these compounds. Molecular simulation offers a safe, efficient means of obtaining such data, and also allows researchers to investigate the chemical and physical properties of molecular systems. The work conducted in this project showed that while Monte Carlo simulations are capable of providing accurate predictions for polar compounds in certain areas, for example, the saturated densities, the corresponding saturated vapour pressures often deviated from experimental data. A similar trend was observed for the binary VLE data simulations: the vapour and liquid compositions were generally well predicted, but the pressures (or temperatures for isobaric data) frequently differed to a certain extent from measured data. Also, this work showed that certain polar species that exhibit complicated chemical behaviour (such as NO_2) are difficult to parameterize successfully. More complex potential models may be required to achieve accurate predictions of the phase behaviour of these compounds.

Simulation of polar compounds is still in the development stages and much work remains to be done before the simulation results are of sufficient reliability and accuracy to replace laboratory measurement as the primary source of experimental

data. However, it is clear that the field of molecular simulation holds much promise. With the continued improvement in potential model complexity and the simultaneous increase in computing power, molecular simulation should become a viable alternative means of generating phase equilibrium data for polar molecules.

10

Recommendations

The work completed during the current study has highlighted several areas that warrant closer scrutiny and investigation. To this end, this chapter offers a number of recommendations that will hopefully point the way towards improving the accuracy and predictive capabilities of Monte Carlo molecular simulations.

1. A general, accurate ab initio method for determining force field parameters that consistently yields a global minimum is crucial if molecular simulation is ever to become completely predictive in nature. The trial-and-error approach adopted in this project (for lack of a better alternative) does not offer any consistency in the parameters that are determined. Moreover, this technique requires experimental data for comparison during the final fitting process, which somewhat defeats the eventual aim of molecular simulation: to provide accurate phase equilibrium predictions without the need for prior laboratory measurements.

2. The sensitivity analysis performed during this study (refer to Section 8.2.5) revealed that great care must be taken when selecting and using combining rules for calculating unlike interaction parameters. This is an area that is often

neglected by researchers, but it was clearly demonstrated that even small variations in these parameters can significantly impact the simulation results.

3. It is recommended that research aimed at improving the currently available combining rules (particularly for systems involving polar compounds) should be undertaken. Particular emphasis should be placed on developing rules that are physically meaningful and have a solid theoretical foundation, as opposed to the popular Lorentz-Berthelot and geometric mixing rules that are largely empirical. Obviously, a strong argument exists for the use of such combining rules with force fields like TraPPE-UA and NERD, which have been parameterized using the Lorentz-Berthelot mixing rule. However, if molecular simulation is ever to become totally independent of measured experimental data (particularly in terms of prediction of multicomponent phase equilibria), novel and innovative combining rules based on rigorous mathematical models are essential and unavoidable.

4. During the parameterization of nitrogen dioxide (see Section 8.2.3), it became clear that an accurate description of the complex NO_2 chemistry was difficult using the relatively uncomplicated potential model proposed here. The problems encountered were largely attributed to the formation of N_2O_4 dimers in the liquid and vapour phases. It is possible that this crucial dimerization behaviour may be successfully introduced into the model through the relatively simple expedient of increasing the NO_2 partial charges. This would, however, require re-parameterization of the Lennard-Jones 12-6 parameters presented in this work. If the increased partial charges do not have the desired effect on the microstructure of the NO_2 system, it is recommended that Morse bonds be introduced into the model in an effort to simulate the $2NO_2 \leftrightarrow N_2O_4$ dimerization reaction.

5. In addition, an attempt to improve the vapour pressure prediction of NO_2 can be made by altering the molecule aspect ratio. The aspect ratio was discussed briefly in Section 8.2.2.2 of the previous chapter, and more details regarding the aspect ratio and the avenues available for altering this molecular property may

be found in the paper by Kamath et al. (2004). The paper also provides several references that could prove useful in this endeavour, such as Stoll et al. (2001).

In addition to the recommendations given above, this project leads naturally into a number of possible topics for future study:

- The solubility of H_2S and SO_2 in hydrocarbon mixtures (this would be of particular interest to the petrochemical sector),
- Binary VLE systems consisting of carboxylic acids with alkanes and alcohols,
- Ternary mixtures involving carboxylic acids,
- Molecular simulation of long-chain carboxylic acids.

It is acknowledged that some of these areas have already been touched upon in the literature (for example, the solubility of H_2S in hydrocarbons: Nath, 2003). However, there remains widespread scope for research and inquiry into all of these areas.

References

Abrams, D. S.; Prausnitz, J. M. (1975) *AIChE Journal* **21**, pp. 116-127.

Aguado, A; Madden, P.A. (2003) *J. Chem. Phys.* **119**, 7471-7483.

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

Askåsen, A; Johansen, D (2002) *Wolfy User Guide*, Rev. 1; University College of Borås.

Barela, R.; Liwski, G.; Szatylowicz, H. (1995) *Fluid Phase Equilib.* **105**, 119-127.

Bates, D. V. (1972) *Am. Rev. Respir. Dis.* **105**, 1-13.

Bernstein, H. J. (2001) *RasMol v2.7.2.1.1*, Copyright © Herbert J. Bernstein 1998-2001.

Berthelot, D. C. (1898) *Hebd. Séanc. Acad. Sci. Paris* **126**, 1703.

Born, M.; von Karman, T. (1912) *Physik. Z.* **13**, 297-309.

Braker, W.; Mossman, A. L. (1980) *Matheson Gas Data Book*, 6th ed.; Matheson: Lyndhurst, NJ.

Campbell, S. W. (1994) *Fluid Phase Equilib.* **102**, 61-84.

Chen, B.; Siepmann, J. I. (2000) *J. Phys. Chem. B* **104**, 8725-8734.

Chen, B.; Potoff, J. J.; Siepmann, J. I. (2001) *J. Phys. Chem. B* **105**, 3093-3104.

Christian, S. D. (1957) *J. Phys. Chem.* **61**, 1441-1442.

Clifford, S. L.; Ramjugernath, D.; Raal, J. D. (2004) *J. Chem. Eng. Data* **49**, 1189-1192.

Clifford, S. L.; Ramjugernath, D.; Raal, J. D. (2005) *Fluid Phase Equilib.* **237**, 89-99.

Côté, A. S.; Smith, B.; Lindan, P. J. D. (2001) http://www.compsoc.man.ac.uk/~lucky/Democritus; Daresbury Laboratory.

Cotton, F. A.; Wilkinson, G.; Munillo, C. A.; Bochmann, M. (1999) *Advanced Inorganic Chemistry*, 6th ed.; John Wiley & Sons, Inc.: New York.

Daubert, T. E.; Danner, R. P. (1989) *Physical and Thermodynamic Properties of Pure Chemicals*, DIPPR, AIChE: New York.

De Leeuw, S. W.; Perram, J. W.; Smith, E. R. (1980a) *Proc. Roy. Soc. London* **A373**, 27-56.

De Leeuw, S. W.; Perram, J. W.; Smith, E. R. (1980b) *Proc. Roy. Soc. London* **A373**, 57-66.

Delhommelle, J.; Boutin, A.; Fuchs, A. H. (1999) *Mol. Simul.* **22**, 351-368.

Delhommelle, J.; Millié, P.; Fuchs, A. H. (2000) *Mol. Phys.* **98**, 1895-1905.

Delhommelle, J.; Millié, P. (2001) *Mol. Phys.* **99**, 619-625.

Dortmund Data Bank (DDB) Software, purchased 1998.

Duggan, A. (2001) *Beowulf Cluster Computing*, Issue V1.R2.M2; Tesella.

Fairchild, G. A.; Roan, J.; McCarroll, J. (1972) *Arch. Environ. Health* **25**, 174-182.

Fisher, M. E. (1964) *Arch. Ration. Mech. Anal.* **17**, 377-410.

Frenkel, D.; Smit, B. (2002) *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: San Diego.

Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, Jr., J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. (1998) *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA.

Glòwka, S. (1972) *Bull. Acad. Pol. Sci.: Ser. Sci. Chim.* **20**, 163-167.

Gmehling, J.; Onken, U. (1977) *Vapour-Liquid Equilibrium Data Collection. DECHEMA Data Series*, Vol. I; DECHEMA: Frankfurt.

Gmehling, J.; Onken, U. (1982) *Vapour-Liquid Equilibrium Data Collection. DECHEMA Data Series*, Vol. VI; DECHEMA: Frankfurt.

Goldstein, E.; Carroll, M. E.; Hoeprich, P. D. (1973) *Arch. Environ. Health* **26**, 202-204.

Goodwin, R. D. (1983) *NBS Report No. NBSIR 83-1694*; National Bureau of Standards: Washington DC.

Halgren, T. A. (1992) *J. Am. Chem. Soc.* **114**, 7827-7843.

Hayden, J. G.; O'Connell, J. P. (1975) *Ind. Eng. Chem. Process Design Develop.* **14**, pp. 209-216.

Heyes, D. M. (1981) *J. Chem. Phys.* **74**, 1924-1929.

Hill, T. L. (1956) *Statistical Mechanics: Principles and Selected Applications*; Dover Publications, Inc.: New York.

Hoover, W. G. (1991) *Computational Statistical Mechanics*; Elsevier Science Publishers B. V.: Amsterdam.

Huang, K. (1963) *Statistical Mechanics*; John Wiley & Sons, Inc.: New York.

Hummer, G.; Grønbech-Jensen, N.; Neumann, M. (1998) *J. Chem. Phys.* **109**, 2791-2797.

Hwang, M.-J.; Ni, X.; Waldman, M.; Ewing, C. S.; Hagler, A. T. (1998) *Biopolymers*, **45**, 435-468.

James, F. (1990) *Comp. Phys, Comm.* **60**, 329-344.

James, F. (1994) *Comp. Phys, Comm.* **79**, 111-114.

Jorgensen, W. L. (1986) *J. Phys. Chem.* **90**, 6379-6388.

Joseph, M.; Raal, J. D.; Ramjugernath, D. (2001) *Fluid Phase Equilib.* **182**, 157-176.

Kamath, G.; Cao, F.; Potoff, J. J. (2004) *J. Phys. Chem. B* **108**, 14130-14136.

Kamath, G.; Lubna, N.; Potoff, J. J. (2005) *J. Chem. Phys.* **123**, 124505.

Kato, M.; Yoshikawa, H.; Yamaguchi, M. (1990) *Fluid Phase Equilib.* **54**, 47-56.

Kohl, A. L.; Riesenfeld, F. C. (1979) *Gas Purification*, 3rd ed.; Gulf Publishing Company: Houston.

Kong, C. L. (1973) *J. Chem. Phys.* **59**, 2464-2467.

Kristóf, T.; Liszi, J. (1997) *J. Phys. Chem. B* **101**, 5480-5483.

Kroschwitz, J. I.; Howe-Grant, M.; Humphreys, L.; Altien, L.; Lee, J.; Vilardo, B. A. (1997) *Encyclopedia of Chemical Technology*, 4th ed., Vol 23; John Wiley & Sons: New York.

Landsberg, H. E. (1981) *The Urban Climate;* Academic Press: New York.

Lebowitz, J. L.; Percus, J. K.; Verlet, L. (1967) *Phys. Rev.* **153,** 250-254.

Leu, A.-D.; Carroll, J. J.; Robinson, D. B. (1992) *Fluid Phase Equilib.* **72,** 163-172.

Lisk, D.J. (1988) *Sci. Total Environ.* **74,** 39-66.

Lorentz, H. A. (1881) *Ann. Physik* **12,** 127.

Lüscher, M. (1994) *Comp. Phys, Comm.* **79,** 100-110.

Marek, J. (1955) *Coll. Czech. Chem. Communications* **20,** 1490-1502.

Marek, J.; Standart, G. (1954) *Coll. Czech. Chem. Communications* **19,** 1074-1084.

Martin, M. G.; Siepmann, J. I. (1998) *J. Phys. Chem. B* **102,** 2569-2577.

Martin, M. G.; Siepmann, J. I. (1999) *J. Phys. Chem. B* **103,** 4508-4517.

Martin, M. G.; Chen, B.; Siepmann, J. I. (2000) *J. Phys. Chem. B* **104,** 2415-2423.

Martin, M. G.; Chen, B.; Wick, C.; Stubbs, J. M.; Potoff, J. J.; Siepmann, J. I. (2005) *MCCCS-Towhee,* `http://towhee.sourceforge.net`.

McClellan, A. L. (1963-1974) *Tables of Experimental Dipole Moments;* W. H. Freeman: San Francisco.

McKnight, T. J.; Vlugt, T. J. H.; Ramjugernath, D.; Starzak, M.; Ahlström, P.; Bolton, K. (2005) *Fluid Phase Equilib.* **232,** 136-148.

Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. (1953) *J. Chem. Phys.* **21,** 1087-1092.

Mooij, G. C. A. M.; Frenkel, D.; Smit, B. (1992) *J. Phys.: Condens. Matter* **4**, L255-L259.

Morachevsky, A. G.; Zharov, V. T. (1963) *Zh. Prikl. Khim.* **36**, 2771 (as reported by Gmehling, J.; Onken, U. (1977) *Vapour-Liquid Equilibrium Data Collection. Organic Hydroxy Compounds: Alcohols*, Vol. I, Part 2a; DECHEMA: Frankfurt).

Morse, P. M. (1929) *Phys. Rev.* **34**, 57-64.

Nath, S. K. (2003) *J. Phys. Chem. B* **107**, 9498-9504.

Nitta, T.; Itami, J.; Katayama, T. (1973) *J. Chem. Eng. Japan* **6**, 303-309.

Nitta, T.; Kido, O.; Katayama, T. (1976) *J. Chem. Eng. Japan* **9**, 317-318.

Nothnagel, K-H.; Abrams, D. S.; Prausnitz, J. M. (1973) *Ind. Eng. Chem. Process Design Develop.* **12**, 25-35.

NPACI (2002) *Rocks Cluster Distribution User Guide v3.1.0*, UC Regents.

Ohe S. (1990) *Vapor-liquid Equilibrium Data at High Pressure;* Elsevier: Amsterdam.

Panagiotopoulos, A. Z. (1987) *Mol. Phys.* **61**, 813-826.

Panagiotopoulos, A. Z.; Quirke, N.; Stapleton, M.; Tildesley, D. J. (1988) *Mol. Phys.* **63**, 527-545.

Panagiotopoulos, A. Z. (1992) *Mol. Simul.* **9**, 1-23.

Pearlman, M. E.; Finklea, J. F.; Creason, J. P.; Shy, C. M.; Young, M. M.; Horton, J. M. (1971) *Pedia.* **47**, 391-398.

Perry, R. H.; Green, D. W.; Maloney, J. O. (1997) *Perry's Chemical Engineers Handbook*, 7th ed.; McGraw-Hill Book Co.: Singapore.

Pitzer, K. S.; Curl, R. F. (1957) *J. Am. Chem. Soc.* **79**, pp. 2369-2370.

Potoff, J. J.; Errington, J. R.; Panagiotopoulos, A. Z. (1999) *Mol. Phys.* **97**, 1073-1083.

Prausnitz, J. M.; Anderson, T. F.; Grens, E. A.; Eckert, C. A.; Hseih, R.; O' Connell, J. P. (1980) *Computer Calculations for Multicomponent Vapour-Liquid and Liquid-Liquid Equilibria*; Prentice-Hall: Englewood Cliffs, NJ.

Prausnitz, J. M.; Lichtenthaler, R. N.; de Azevedo, E. G. (1999) *Molecular Thermodynamics of Fluid-Phase Equilibria*, 3rd ed.; Prentice-Hall: Englewood Cliffs, NJ.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. (1986) *Numerical Recipes: The Art of Scientific Computing*; Cambridge University Press: Cambridge.

Radajewski, J.; Eadline, D. (1998) *Beowulf HOWTO*, v1.1.1, Linux Documentation Project.

Reamer, H. H.; Sage, B. H.; Lacey, W. N. (1951) *Ind. Eng. Chem.* **43**, 976-980.

Reamer, H. H.; Sage, B. H. (1952) *Ind. Eng. Chem.* **44**, 185-187.

Reid, C. R.; Prausnitz, J. M.; Polling, B. E. (1977) *The Properties of Gases and Liquids*, 3rd ed.; McGraw-Hill Book Company: Singapore.

Reid, C. R.; Prausnitz, J. M.; Polling, B. E. (1988) *The Properties of Gases and Liquids*, 4th ed.; McGraw-Hill Book Company: Singapore.

Renon, H.; Prausnitz, J. M. (1968) *AIChE Journal* **14**, pp. 135-144.

Rosenbluth, M. N.; Rosenbluth, A. W. (1955) *J. Chem. Phys.* **23**, pp. 356-359.

Rowlinson, J. S.; Swinton, F. L. (1982) *Liquids and Liquid Mixtures*, 3rd ed.; Butterworth: London.

Sadus, R. J. (1993) *J. Phys. Chem.* **97**, 1985-1992.

Sebastiani, E.; Lacquaniti, L. (1967) *Chem. Eng. Sci.* **22**, 1155-1162.

Sharma, R.; Baid, A.; Jayaram, I. S. (1999) *Ind. Eng. Chem. Res.* **38**, 1057-1064.

Sharma, R.; Yesavage, V. F.; Kidnay, A. J. (1983) *AIChE Journal* **29**, 508-511.

Shih, T. T.; Jones, D. K. (1989) *Experimental Results from the Design Institute for Physical Property Data: Phase Equilibria and Pure Component Properties Part II*, AIChE Symposium Series No. 271, Vol. 85, AIChE.

Siepmann, J. I.; Frenkel, D. (1992) *Mol. Phys.* **70**, pp. 1145-1158.

Smit, B.; de Smedt, P.; Frenkel, D. (1989) *Mol. Phys.* **68**, 931-950.

Smit, B.; Karaborni, S.; Siepmann, J. I. (1995) *J. Chem. Phys.* **102**, 2126-2140. Erratum: (1998) *J. Chem. Phys.* **109**, 352.

Smit, B.; Frenkel, D. (1989) *Mol. Phys.* **68**, 951-958.

Smith, B. D.; Srivastava, R. (1986) *Thermodynamic Data for Pure Compounds: Part B Halogenated Hydrocarbons and Alcohols*; Elsevier: Amsterdam.

Smith, R. L. (1974) *Ecology and Field Biology*, 2nd ed.; Harper & Row: New York.

Smith, J. M.; Van Ness, H. C.; Abbott, M. M. (1996) *Introduction to Chemical Engineering Thermodynamics*, 5th ed.; McGraw-Hill International Editions: New York.

Speight, J. G. (1998) *The Chemistry and Technology of Petroleum*, 3rd ed.; Marcel Dekker, Inc.: New York.

Stoll, J.; Vrabec, J.; Hasse, H.; Fischer, J. (2001) *Fluid Phase Equilib.* **179**, 339-362.

Tamir, A.; Wisniak, J. (1975) *Chem. Eng. Sci.* **30,** 335-342.

Tamir, A.; Wisniak, J. (1978) *Chem. Eng. Sci.* **33,** 651-656.

Tolman, R. C. (1979) *The Principles of Statistical Mechanics;* Dover Publications, Inc.: New York.

Tsonopoulos, C.; Prausnitz, J. M. (1970) *Chem. Eng. J.* **1,** 273-278.

Tsuboka, T.; Katayama, T. (1975) *J. Chem. Eng. Japan* **8,** 181-187.

Twu, C. H.; Coon, J. E. (1996) *AIChE Journal* **42,** 3212-3222.

Vargaftik, N. B. (1975) *Tables on the Thermophysical Properties of Liquids and Gases,* 2nd ed.; Hemisphere Publishing Corporation: Washington.

Vlugt, T. J. H.; Martin, M. G.; Smit, B.; Siepmann, J. I.; Krishna, R. (1998) *Mol. Phys.* **94,** 727-733.

Vlugt, T. J. H.; Krishna, R.; Smit, B. (1999) *J. Phys. Chem. B* **103,** pp. 1102-1118.

Walas, S. M. (1985) *Phase Equilibria in Chemical Engineering;* Butterworth Publishers: Boston.

Waldman, M.; Hagler, A. T. (1993) *J. Comput. Chem.* **14,** 1077-1084.

Welsh, M.; Dalheimer, M. K.; Dawson, T.; Kaufman, L. (2003) *Running Linux,* 4th ed.; O'Reilly and Associates, Inc.: Sebastopol, CA.

Widom, B. (1963) *J. Chem. Phys.* **39,** 2802-2812.

Wilson, G. M. (1964) *J. Am. Chem. Soc.* **86,** pp. 127-130.

Wong, D. S.; Sandler, S. I. (1992) *AIChE Journal* **38,** No.5, 671-680.

A

Pure Component Properties

The pure component properties given here are for the carboxylic acids modeled using chemical theory (refer to Chapter 2 and Chapter 8, Section 8.1). The critical pressure, temperature, volume and compressibility were easily obtained from the Dortmund Data Bank (1998). The dipole moments were found in the compilation of McClellan (1974) and the text of Reid et al. (1988), whilst the solvation and association parameters were acquired from Prausnitz et al. (1980).

The mean radius of gyration was determined using the group contribution method proposed by Reid et al. (1977). The acentric factors were determined from the vapour pressure data measured during the Masters work (Clifford, 2004) involving the carboxylic acids. The Antoine equation coefficients necessary for calculating saturated pressures from Equation 2.16 were calculated by fitting to the vapour pressure data of Clifford et al. (2004). In order to determine the chemical equilibrium constants (K) from Equation 8.1, the constants A_K and B_K were required. These were obtained from the Design Institute for Physical Properties (DIPPR) databank (Shih and Jones, 1989).

Table A.1: Pure component properties for the carboxylic acids modeled using chemical theory.

Pure Component Property	Propanoic Acid	2-methyl propanoic Acid	Pentanoic Acid
T_c (K)	604.0	605.0	643.0
P_c (atm)	44.70	36.50	35.82
V_c (cm^3/mol)	230.0	292.0	340.0
Z_c	0.2003	0.2148	0.2253
Dipole Moment (debye)	1.76	1.34	1.81
Solvation & Association	4.5	4.5	4.5
Radius of Gyration (Å)	2.682	3.115	3.541
Acentric Factor	0.5397	0.6321	0.7165
A_K	-10.483	-9.039	-8.267
B_K	3592.3	2977.0	2702.0
A'	18.1057	15.1762	36.4104
B'	5640.34	3527.86	30029.23
C'	277.46	180.51	760.45

B

TraPPE-UA Parameters

The TraPPE-UA potential model was used for all the simulations carried out during this project, except (obviously) for the pure component simulations involving H_2S, SO_2 and NO_2. The TraPPE-UA parameters may be obtained from the published papers; however, it was decided to group all the parameters relevant to this study in an appendix for reader convenience. The Lorentz-Berthelot combining rules are used to determine unlike interaction parameters for the TraPPE-UA force field.

Table B.1: TraPPE-UA bond lengths (the TraPPE-UA force field uses fixed bond lengths).

Bond	Length (Å)
CH_x–CH_y	1.540
CH_x–OH	1.430
O–H	0.945
C=O	1.214
C–O	1.364
C–CH_x	1.520
O–H (acid)	0.970

Table B.2: TraPPE-UA angle bending parameters.

Bend	Angle (°)	k_θ/k_B (K.rad^{-2})
CH_x-CH_2-CH_y	114	62500
CH_x-CH-CH_y	112	62500
CH_x-C-CH_y	109.47	62500
CH_x-CH_y-O	109.47	50400
CH_x-O-H	108.5	55400
C-O-H	107	17600
O=C-O	123	40300
CH_x-C-O	111	35300
CH_x-C=O	126	40300

Table B.3: TraPPE-UA dihedral angle (torsional) parameters.

Dihedral	c_0/k_B (K)	c_1/k_B (K)	c_2/k_B (K)	c_3/k_B (K)	f_1 (°)
CH_x-CH_2-CH_2-CH_y	0	335.03	-68.19	791.32	0
CH_x-CH_2-CH_2-OH	0	176.62	-53.34	769.93	0
CH_x-CH_2-O-H	0	209.82	-29.17	187.93	0
CH_x-CH-O-H	215.96	197.33	31.46	-173.92	0
O=C-O-H	0	630.0	1562.4	0	180
CH_x-C-O-H	0	630.0	1562.4	0	0
CH_x-CH_2-C=O	0	630.0	1562.4	0	180
CH_x-CH-C=O	0	630.0	1562.4	0	180
		$V1$ (K)	$V2$ (K)	$V3$ (K)	
CH_x-CH_2-C-O		710.06	-136.38	1582.64	
CH_x-CH-C-O		710.06	-136.38	1582.64	
CH_x-CH_2-CH_2-CH_x		710.06	-136.38	1582.64	

The torsional potential used in the TraPPE-UA force field takes the following form for the alkanes and alcohols:

$$U_{torsion} = c_0 + c_1[1+\cos(\phi)] + c_2[1-\cos(2\phi)] + c_3[1+\cos(3\phi)] \quad \text{(B.1)}$$

and

$$U_{torsion} = c_1[1+\cos(\phi+f_1)] + c_2[1-\cos^2\phi] \quad \text{(B.2)}$$

for the carboxylic acids (Martin et al., 2005), with the dihedral angles for the alkane tail of pentanoic acid controlled by the OPLS-UA torsional potential (Kamath et al., 2004):

$$U_{torsion} = \frac{V_1}{2}(1+\cos\phi) + \frac{V_2}{2}(1-\cos(2\phi)) + \frac{V_3}{2}(1+\cos(3\phi)) \quad \text{(B.3)}$$

Table B.4: TraPPE-UA non-bonded parameters.

Pseudo-atom	ε_{ii}/k_B (K)	σ_{ii} (Å)	q_i (e)
CH_4	148.0	3.73	0.000
CH_3-(CH_x)	98.0	3.75	0.000
CH_3-(OH)	98.0	3.75	0.265
CH_3-(C=)	98.0	3.75	0.120
(CH_x)-CH_2-(CH_x)	46.0	3.95	0.000
CH_2-(OH)	46.0	3.95	0.265
CH_2-(C=)	46.0	3.95	0.120
(CH_x)$_2$-CH-(OH)	10.0	4.33	0.265
(CH_x)$_3$-C-(OH)	0.5	5.80	0.265
(O)=C	41.0	3.90	0.420
(CH_x)-O-(H)	93.0	3.02	-0.700
(O)=(C)-O-(H)	93.0	3.02	-0.460
O=(C)	79.0	3.05	-0.450
(CH_x)-(O)-H	0.0	0.00	0.435
(O)=(C)-(O)-H	0.0	0.00	0.370

The parameters presented in Tables B.1 to B.4 were taken from Martin and Siepmann (1998), Chen et al. (2001) and Kamath et al. (2004).

C

Additional Figures

C.1 Activity Coefficient Plots

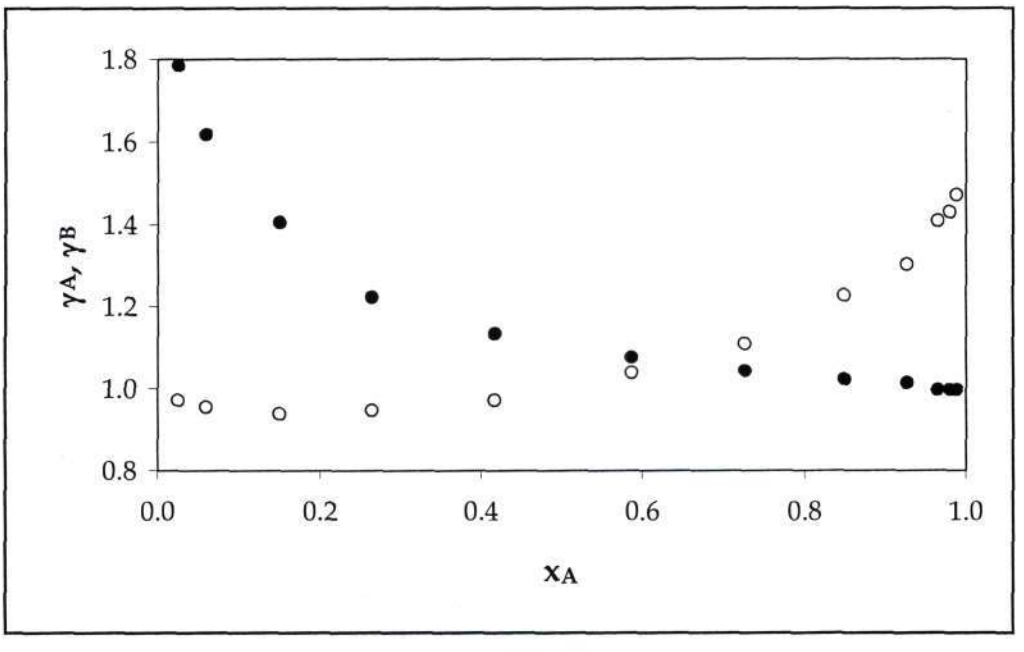

Figure C.1: Liquid-phase activity coefficients, γ_A (filled symbols) and γ_B (open symbols) for the propanoic acid (A) + pentanoic acid (B) system at 393.15 K.

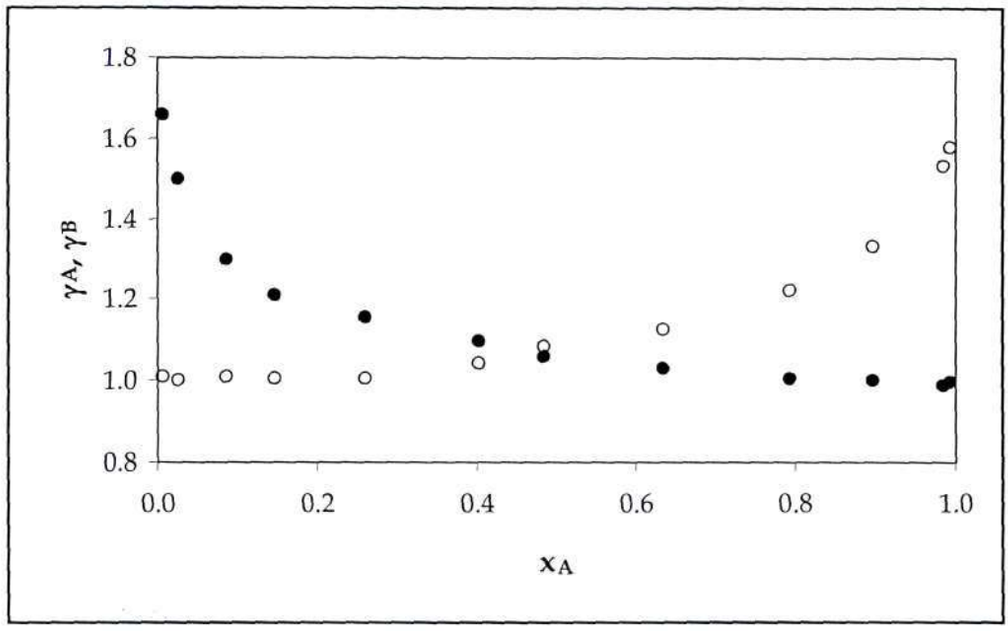

Figure C.2: Liquid-phase activity coefficients, γ_A (filled symbols) and γ_B (open symbols) for the propanoic acid (A) + pentanoic acid (B) system at 413.15 K.

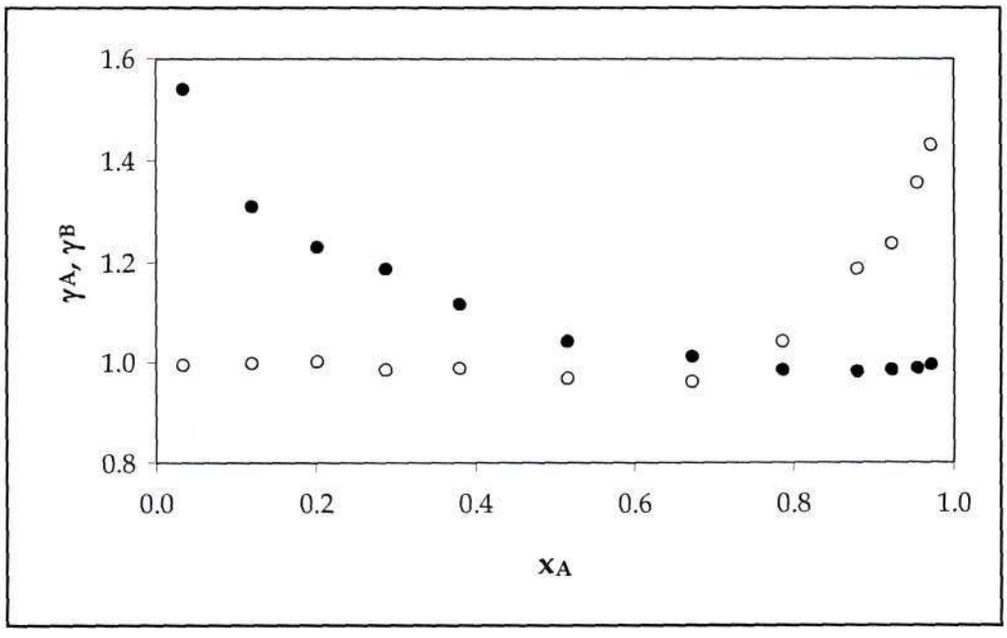

Figure C.3: Liquid-phase activity coefficients, γ_A (filled symbols) and γ_B (open symbols) for the propanoic acid (A) + pentanoic acid (B) system at 20 kPa.

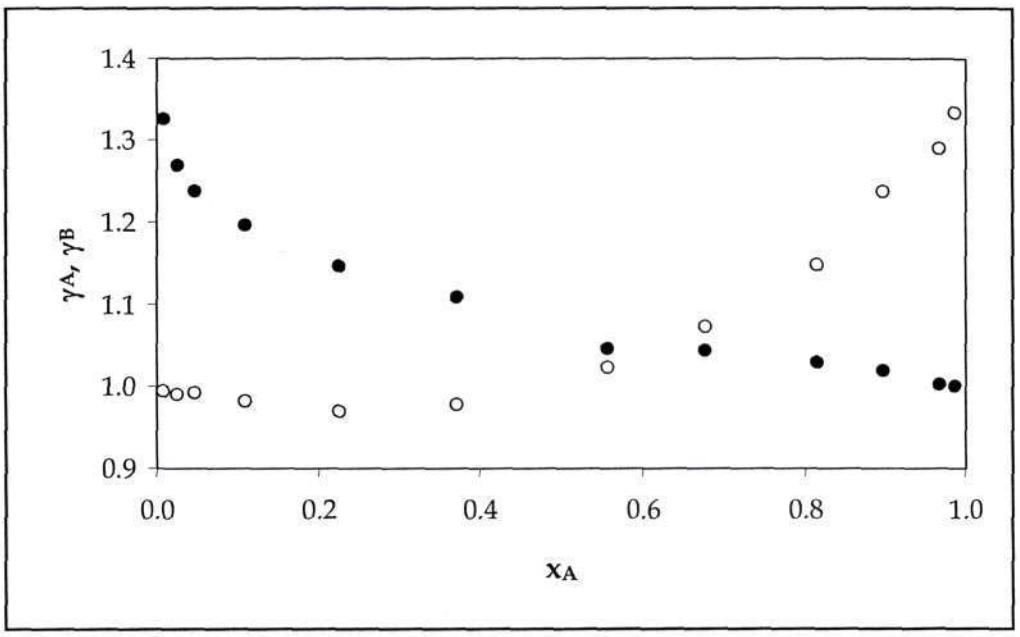

Figure C.4: Liquid-phase activity coefficients, γ_A (filled symbols) and γ_B (open symbols) for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 393.15 K.

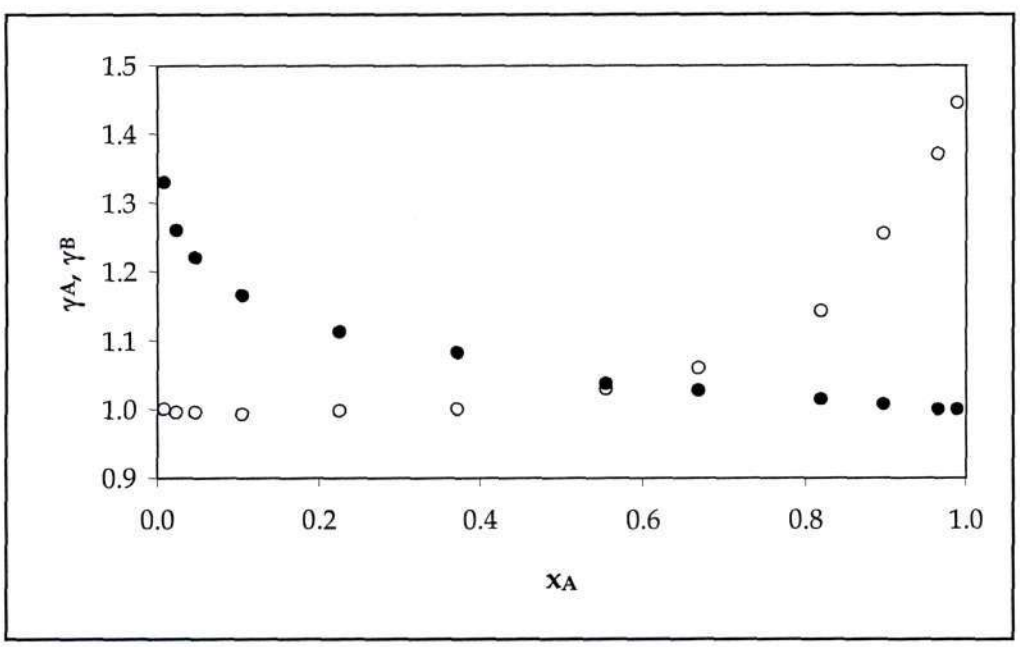

Figure C.5: Liquid-phase activity coefficients, γ_A (filled symbols) and γ_B (open symbols) for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 403.15 K.

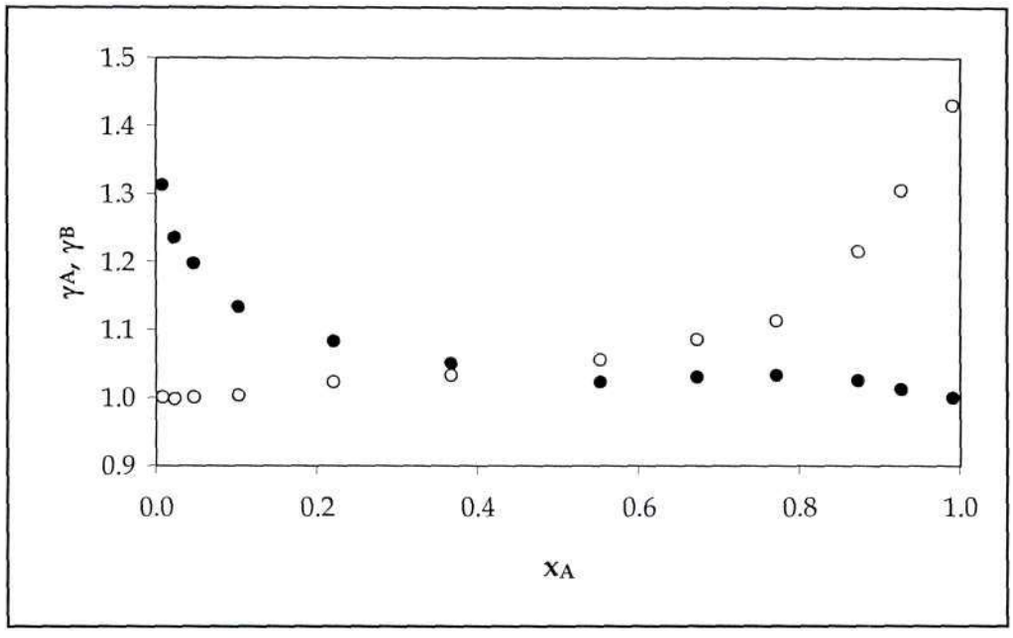

Figure C.6: Liquid-phase activity coefficients, γ_A (filled symbols) and γ_B (open symbols) for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 413.15 K.

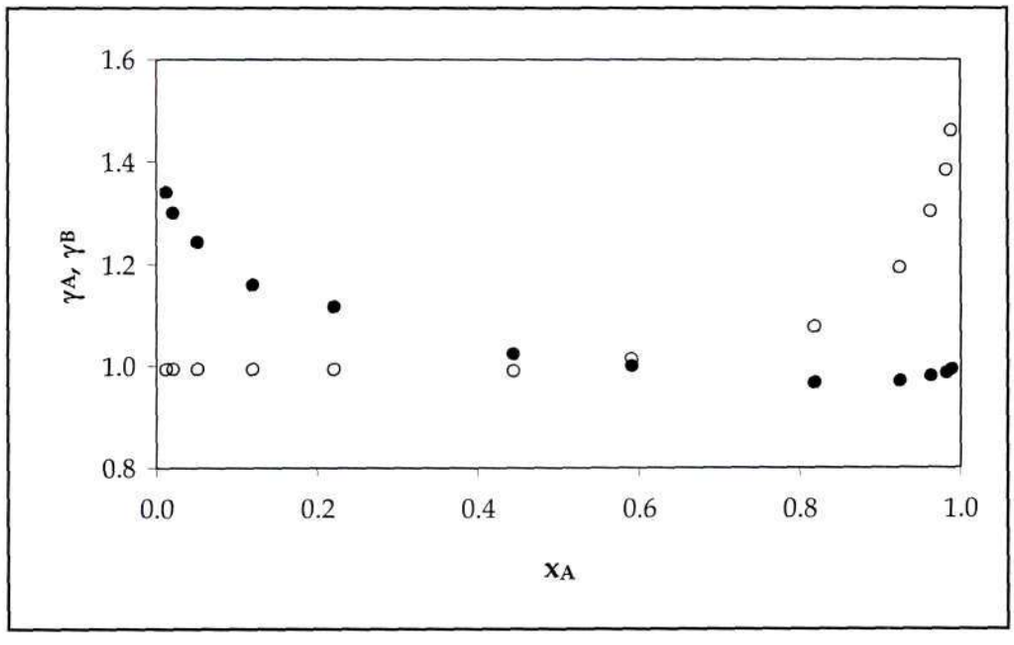

Figure C.7: Liquid-phase activity coefficients, γ_A (filled symbols) and γ_B (open symbols) for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 20 kPa.

C.2 Thermodynamic Consistency Test Plots

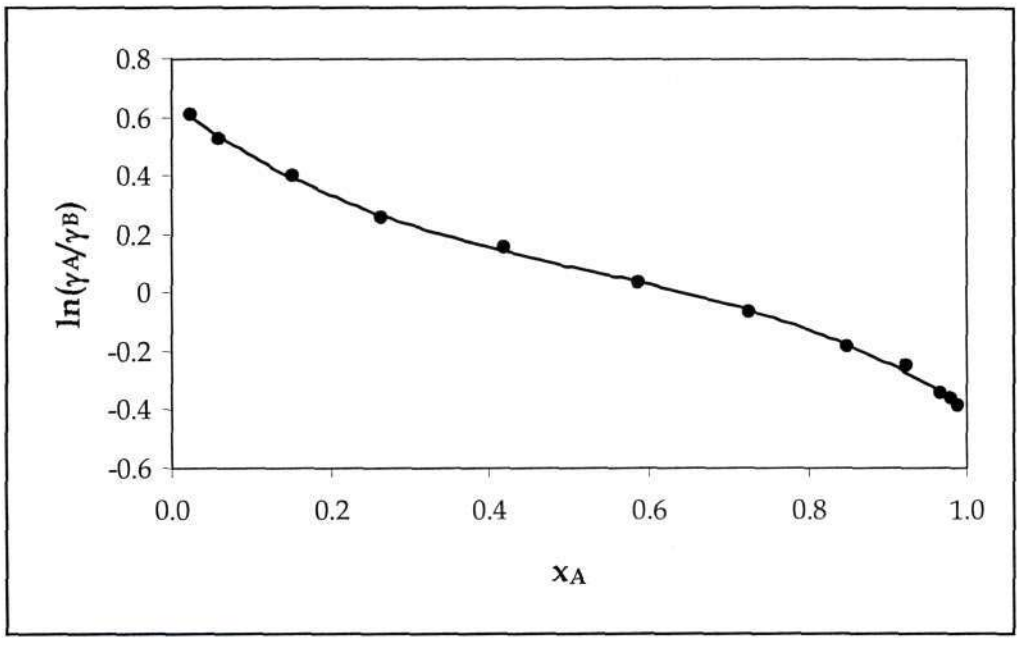

Figure C.8: Area test plot to determine the thermodynamic consistency of the propanoic acid (A) + pentanoic acid (B) system at 393.15 K.

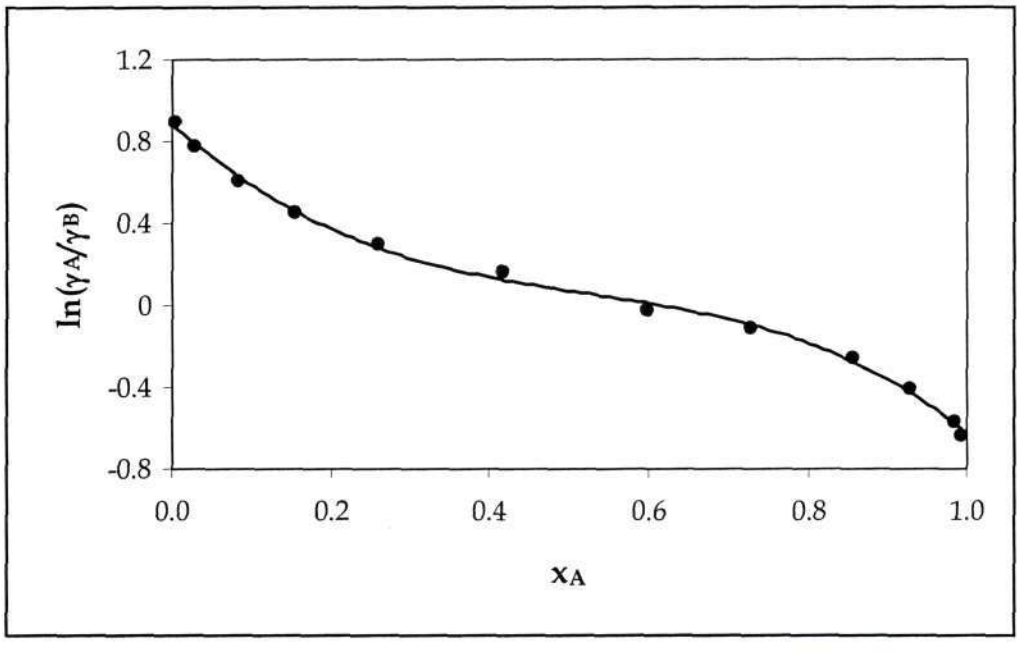

Figure C.9: Area test plot to determine the thermodynamic consistency of the propanoic acid (A) + pentanoic acid (B) system at 403.15 K.

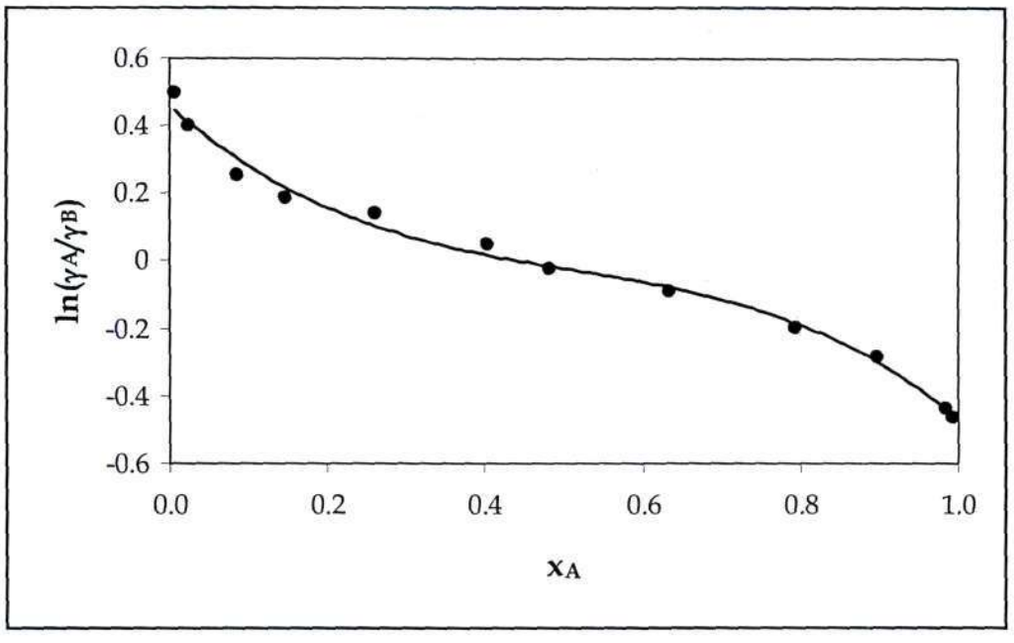

Figure C.10: Area test plot to determine the thermodynamic consistency of the propanoic acid (A) + pentanoic acid (B) system at 413.15 K.

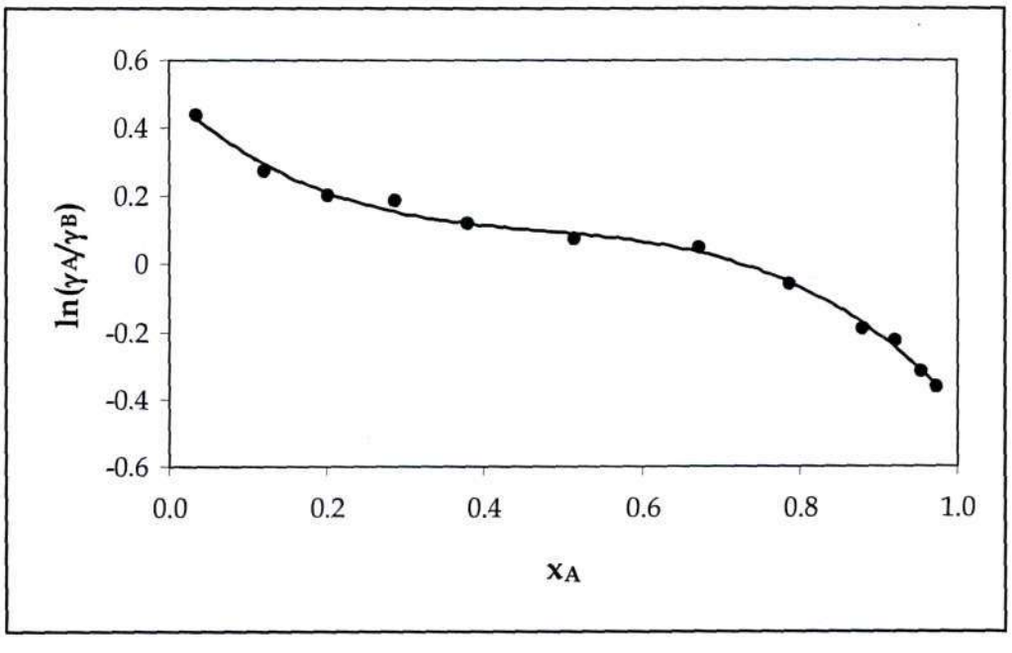

Figure C.11: Area test plot to determine the thermodynamic consistency of the propanoic acid (A) + pentanoic acid (B) system at 20 kPa.

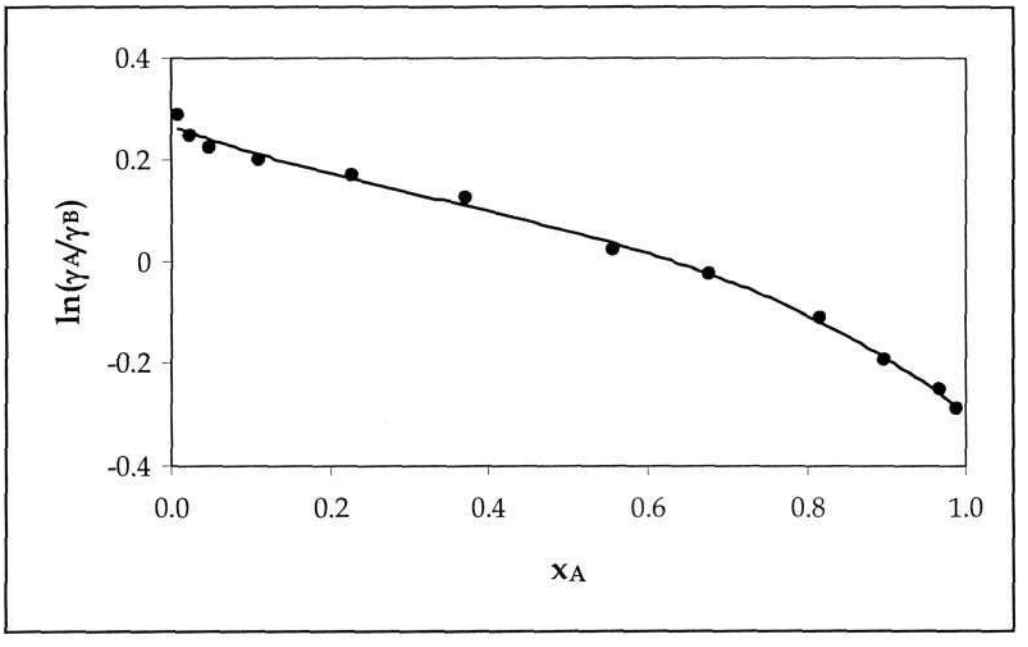

Figure C.12: Area test plot to determine the thermodynamic consistency of the 2- methyl propanoic acid (A) + pentanoic acid (B) system at 393.15 K.

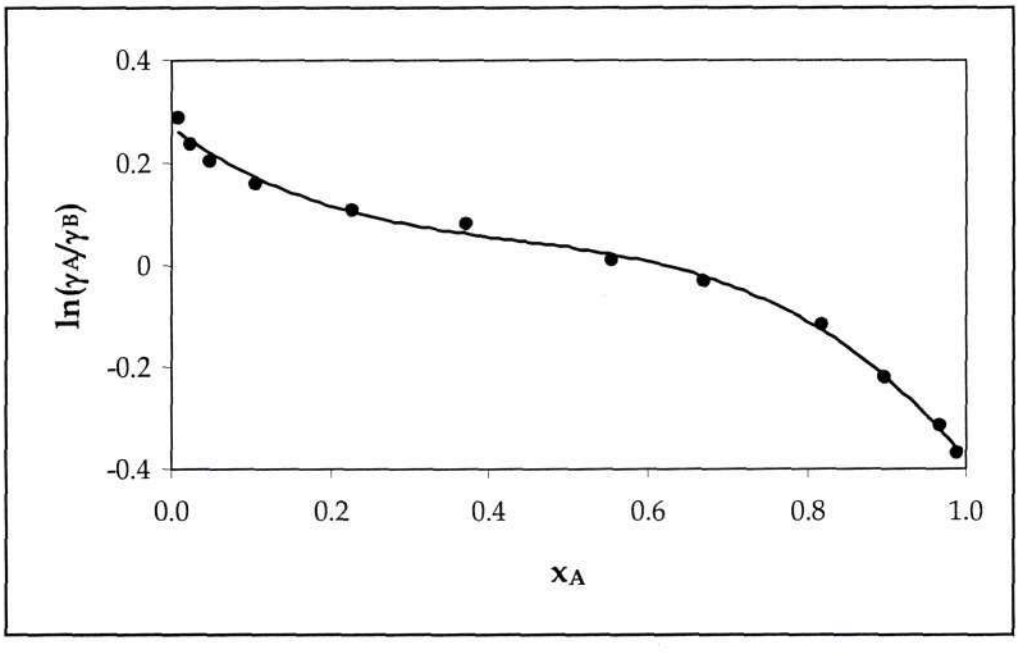

Figure C.13: Area test plot to determine the thermodynamic consistency of the 2- methyl propanoic acid (A) + pentanoic acid (B) system at 403.15 K.

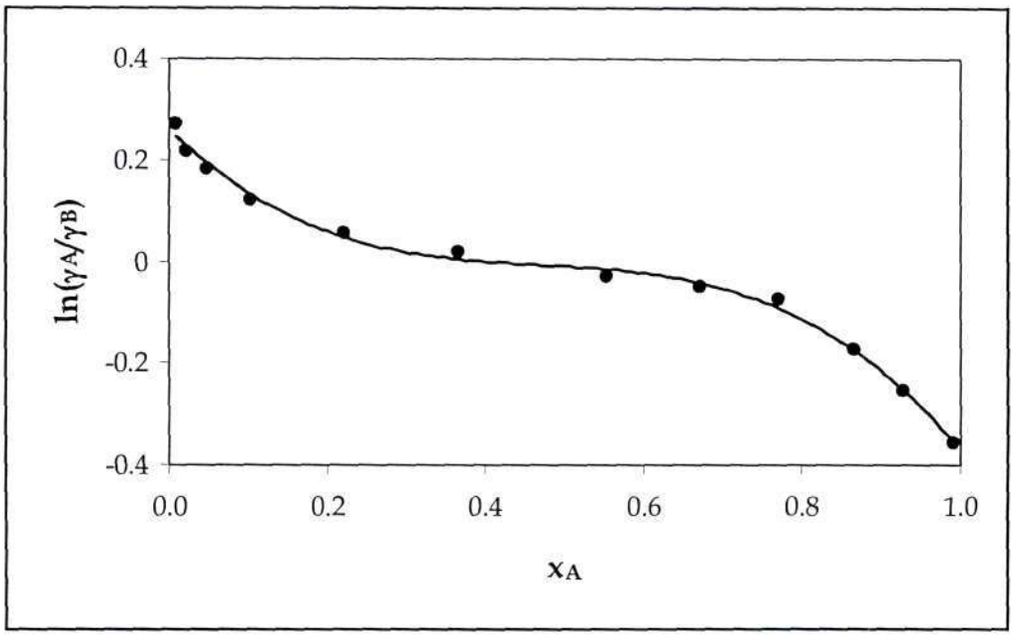

Figure C.14: Area test plot to determine the thermodynamic consistency of the 2- methyl propanoic acid (A) + pentanoic acid (B) system at 413.15 K.

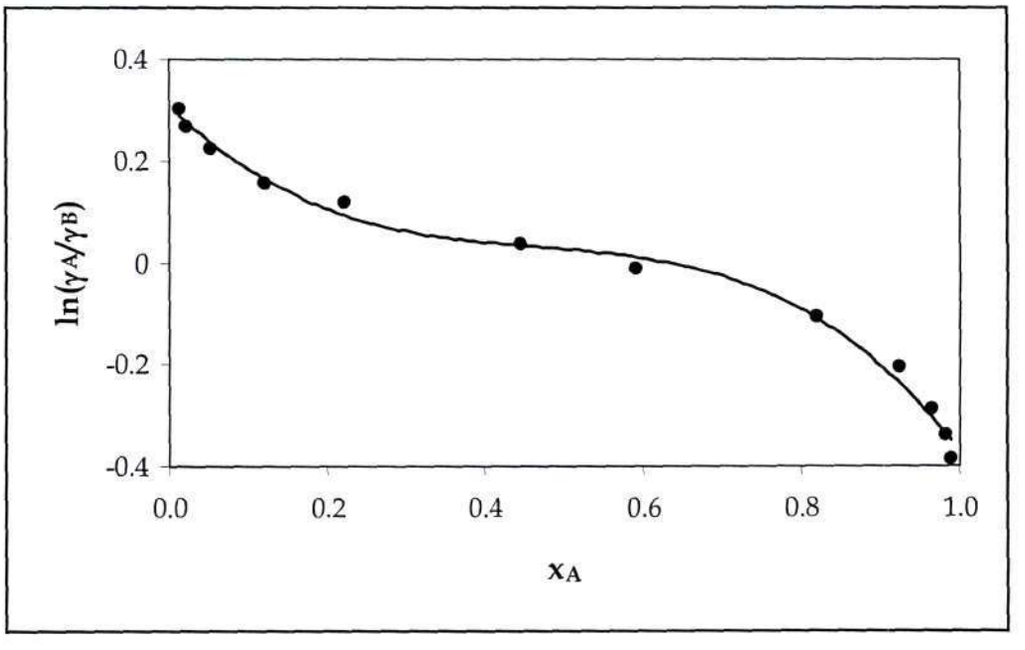

Figure C.15: Area test plot to determine the thermodynamic consistency of the 2- methyl propanoic acid (A) + pentanoic acid (B) system at 20 kPa.

C.3 Chemical Theory Bubble-point Technique Plots

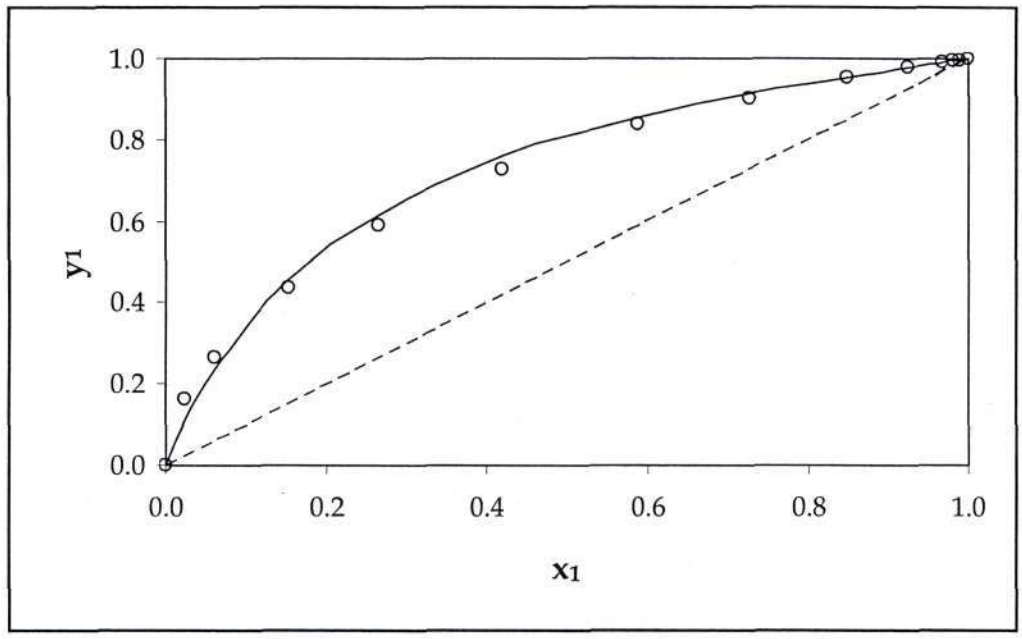

Figure C.16: *x-y* plot for the propanoic acid (A) + pentanoic acid (B) system at 393.15 K. Solid lines are the experimental data and open symbols the model prediction.

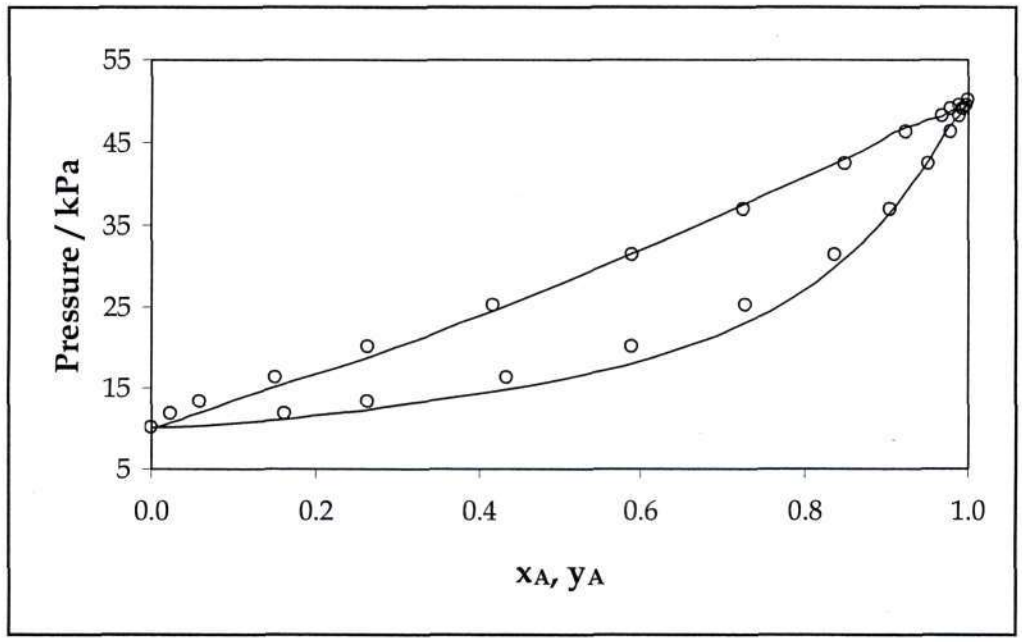

Figure C.17: *P-x-y* plot for the propanoic acid (A) + pentanoic acid (B) system at 393.15 K. Solid lines are the experimental data and open symbols the model prediction.

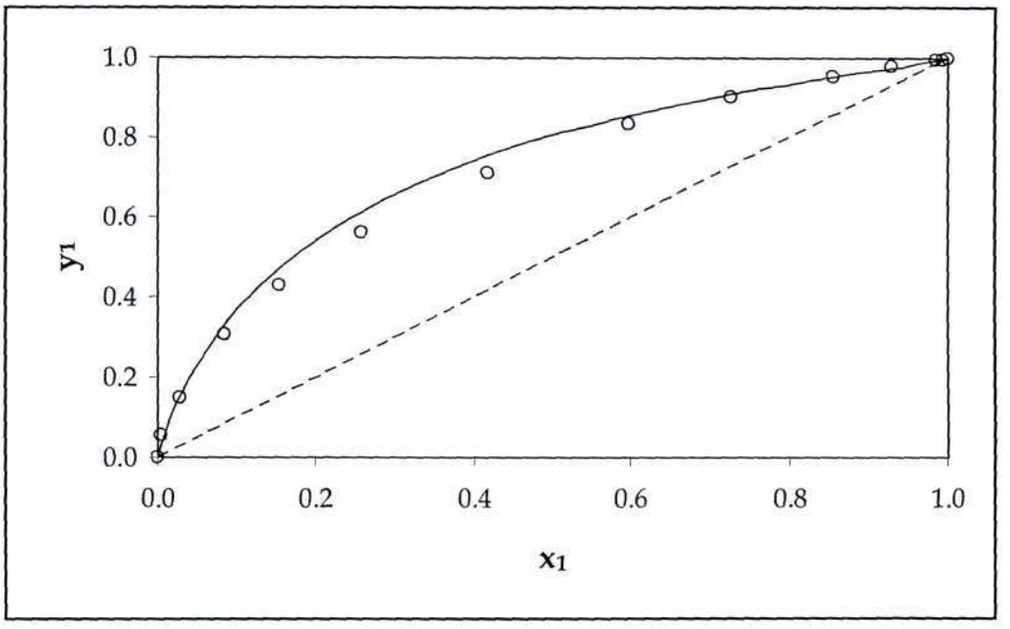

Figure C.18: *x-y* plot for the propanoic acid (A) + pentanoic acid (B) system at 403.15 K. Solid lines are the experimental data and open symbols the model prediction.

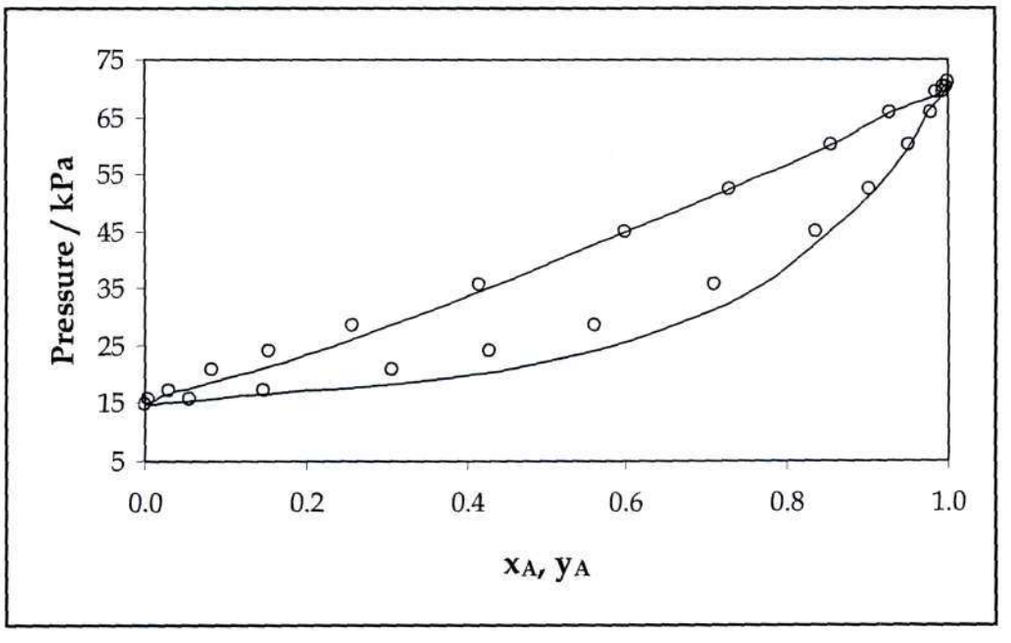

Figure C.19: *P-x-y* plot for the propanoic acid (A) + pentanoic acid (B) system at 403.15 K. Solid lines are the experimental data and open symbols the model prediction.

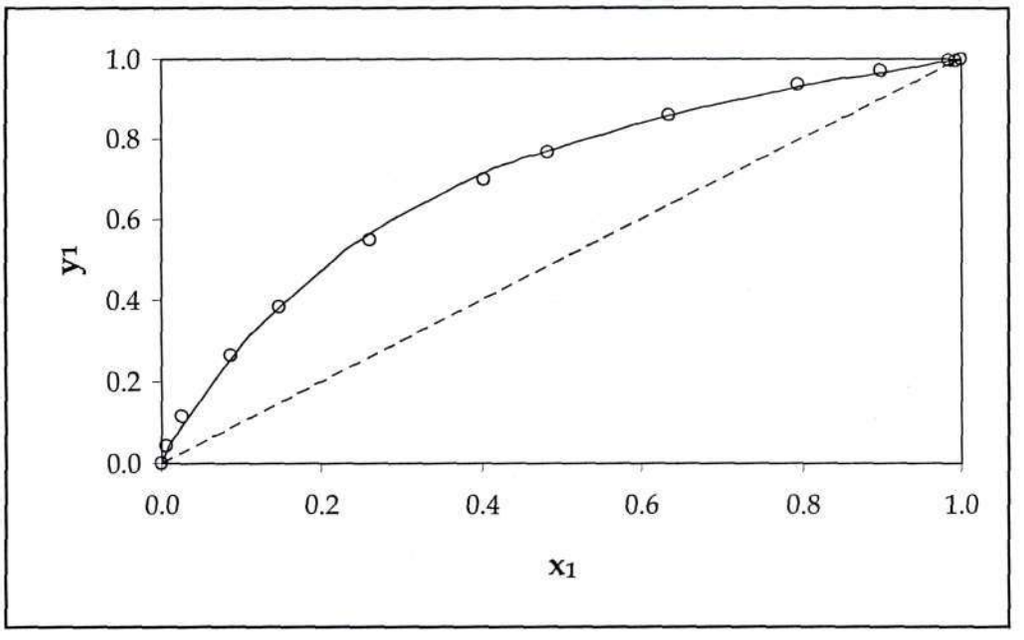

Figure C.20: *x-y* plot for the propanoic acid (A) + pentanoic acid (B) system at 413.15 K. Solid lines are the experimental data and open symbols the model prediction.

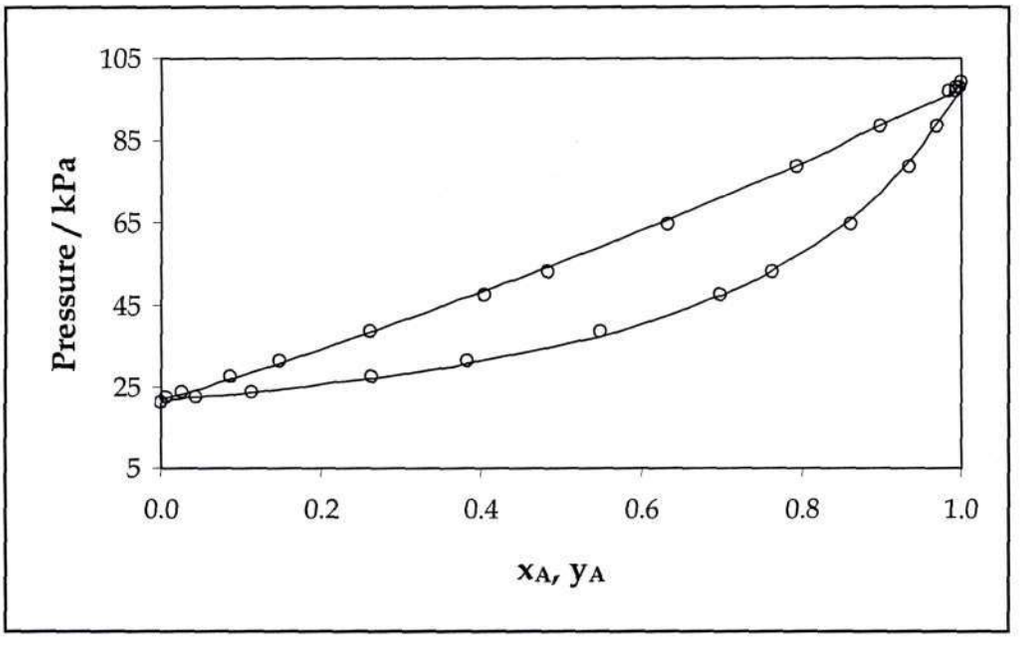

Figure C.21: *P-x-y* plot for the propanoic acid (A) + pentanoic acid (B) system at 413.15 K. Solid lines are the experimental data and open symbols the model prediction.

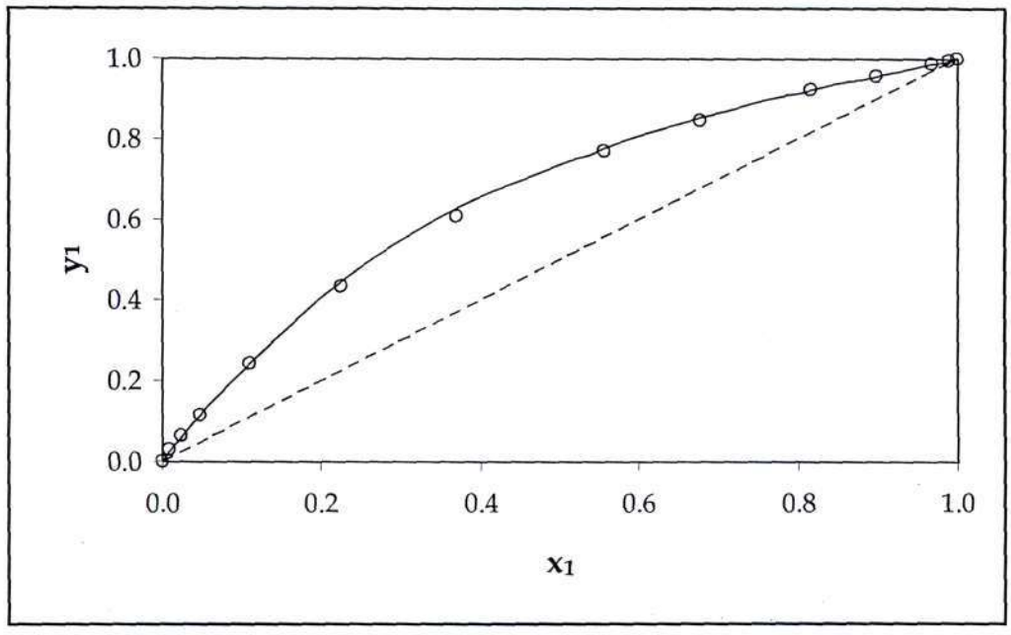

Figure C.22: *x-y* plot for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 393.15 K. Solid lines are the experimental data and open symbols the model prediction.

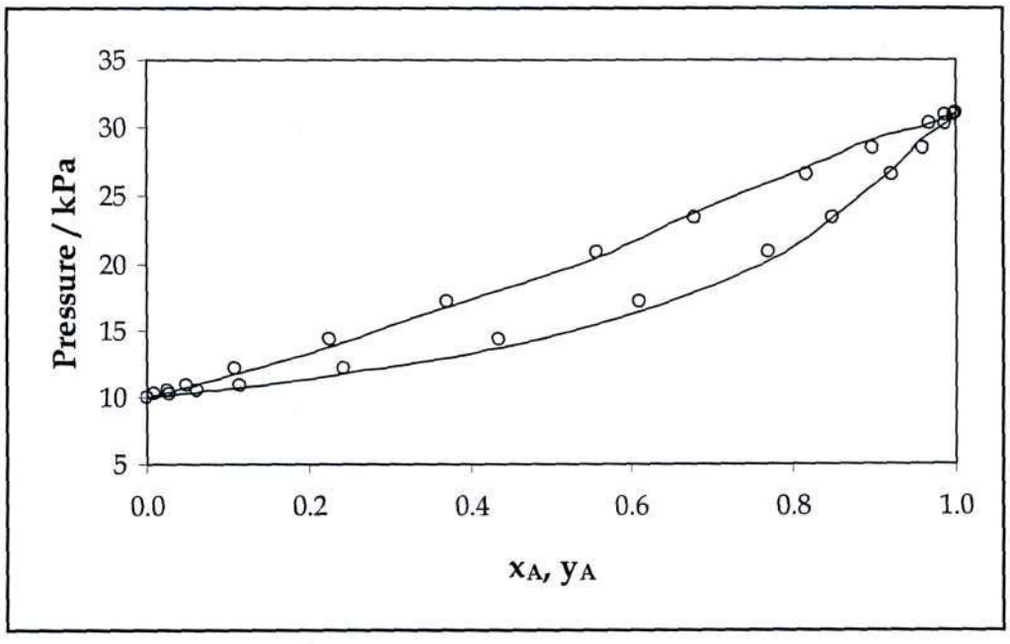

Figure C.23: *P-x-y* plot for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 393.15 K. Solid lines are the experimental data and open symbols the model prediction.

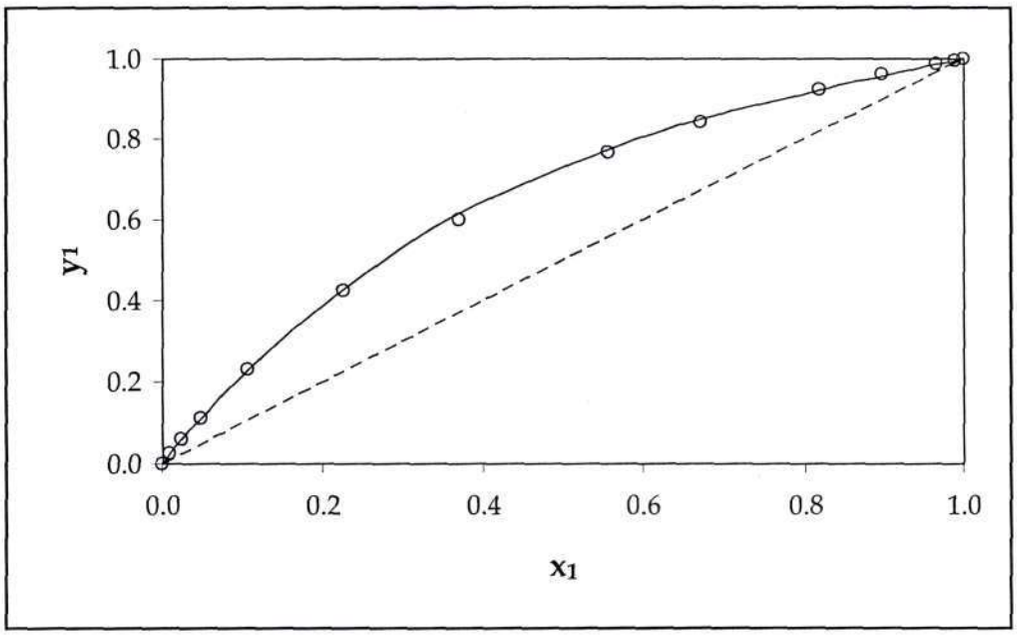

Figure C.24: *x-y* plot for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 403.15 K. Solid lines are the experimental data and open symbols the model prediction.

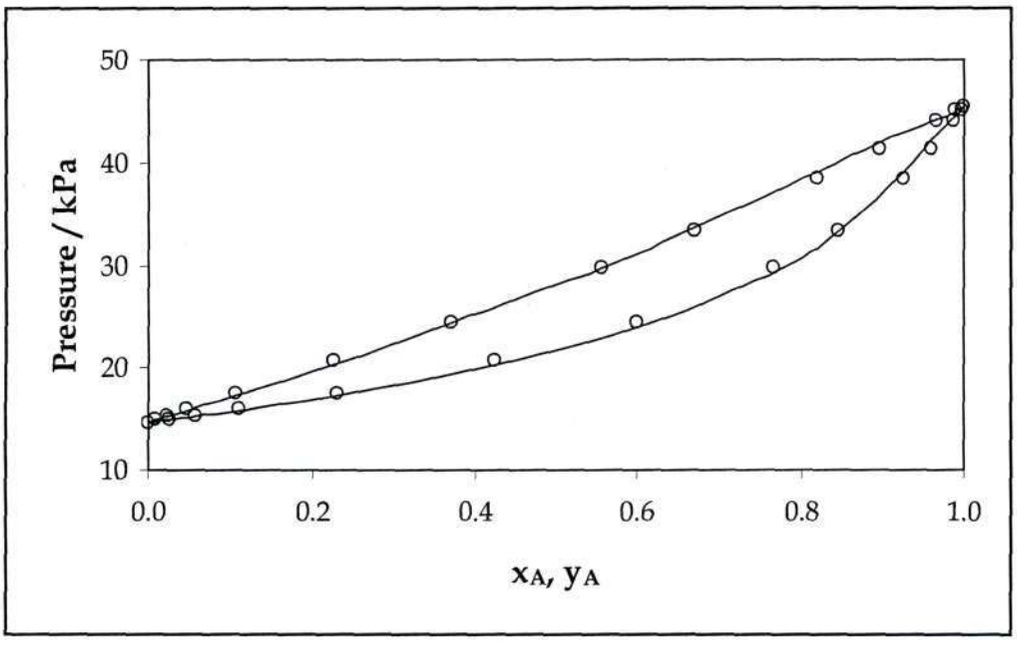

Figure C.25: *P-x-y* plot for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 403.15 K. Solid lines are the experimental data and open symbols the model prediction. (Also given as an example in Section 8.1.3)

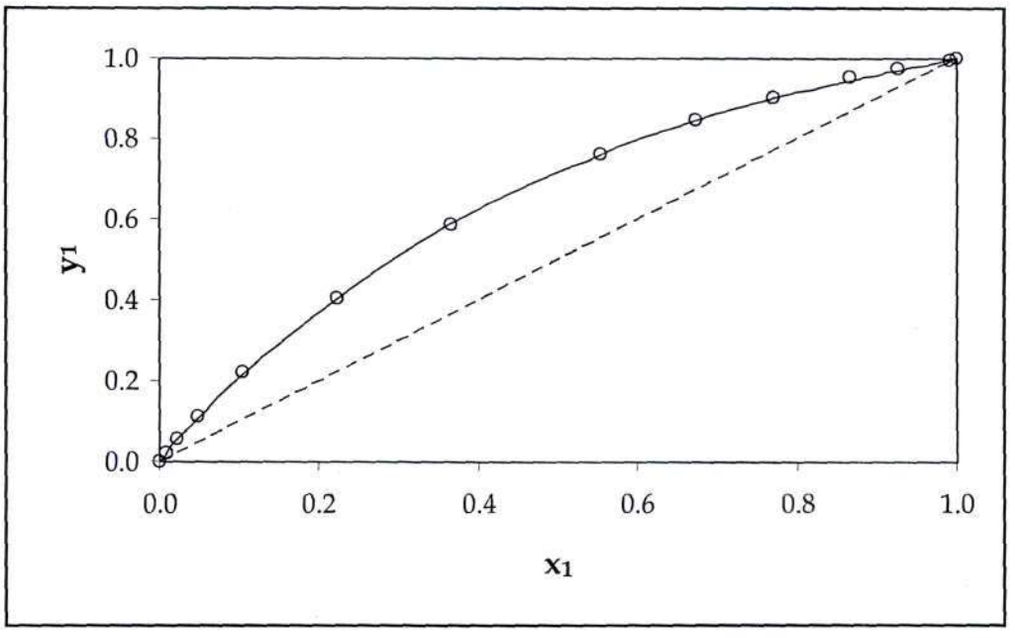

Figure C.26: *x-y* plot for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 413.15 K. Solid lines are the experimental data and open symbols the model prediction.

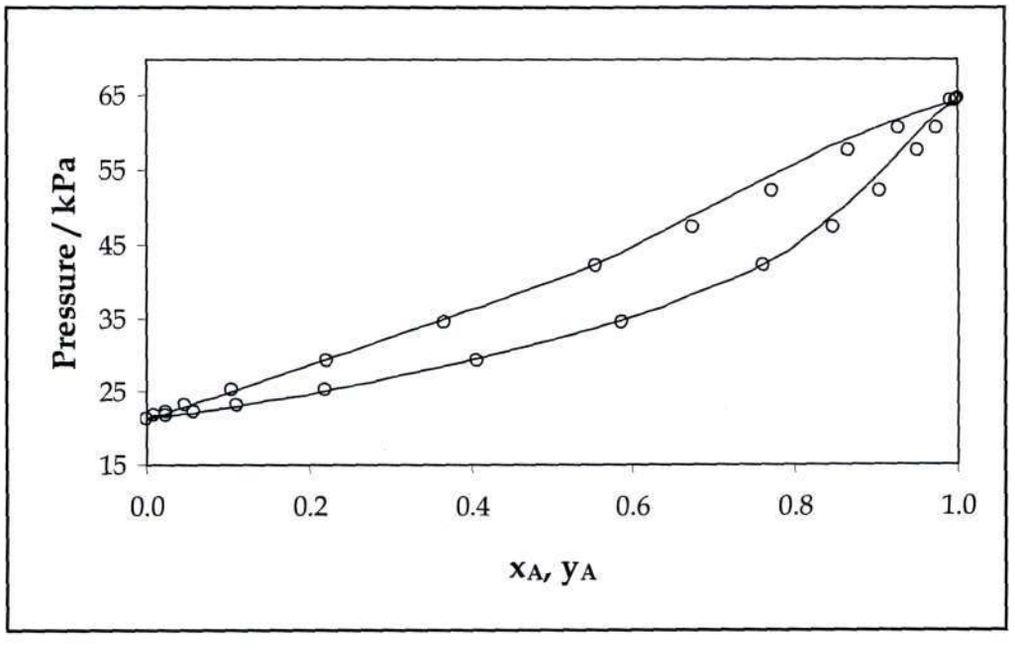

Figure C.27: *P-x-y* plot for the 2-methyl propanoic acid (A) + pentanoic acid (B) system at 413.15 K. Solid lines are the experimental data and open symbols the model prediction.

List of Publications

This section provides a list of the publications arising from the work performed during this project.

- Clifford, S. L.; Ramjugernath, D.; Raal, J. D. (2005) "Vapour-liquid Equilibrium of Carboxylic Acid Systems: Propionic Acid + Valeric Acid and Isobutyric Acid + Valeric Acid", *Fluid Phase Equilib.* **237**, 89-99.

- Clifford, S. L.; Bolton, K.; Ramjugernath, D. (2006) "Monte Carlo Simulation of Carboxylic Acid Phase Equilibria", *J. Phys. Chem. B*, in press.

- Clifford, S. L.; Bolton, K.; Ramjugernath, D. (2006) "Parameterization of H_2S, SO_2 and NO_2", *J. Phys. Chem. B*, submitted for review.

Every day you may make progress. Every step may be fruitful. Yet there will stretch out before you an ever-lengthening, ever-ascending, ever-improving path. You know you will never get to the end of the journey. But this, so far from discouraging, only adds to the joy and glory of the climb.

Sir Winston Churchill

To err is human, but to really foul things up requires a computer.

Farmers' Almanac, 1978