Pro gradu -tutkielma
Teoreettinen fysiikka

Water-ethanol mixtures by molecular dynamics and x-ray Compton scattering

Iina Juurinen
2009

Ohjaaja:	Dos. Mikko Hakala
Tarkastajat:	Prof. Keijo Hämäläinen
	Prof. Kai Nordlund

HELSINGIN YLIOPISTO
FYSIIKAN LAITOS

PL 64 (Gustaf Hällströmin katu 2)
00014 Helsingin yliopisto

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Matemaattis-luonnontieteellinen tiedekunta	Fysiikan laitos

Tekijä — Författare — Author
Iina Juurinen

Työn nimi — Arbetets titel — Title

Vesi-etanoliseosten tutkimus molekyylidynamiikalla ja Compton-sironnalla

Oppiaine — Läroämne — Subject
Teoreettinen fysiikka

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Pro gradu -tutkielma	Syyskuu 2009	51

Tiivistelmä — Referat — Abstract

Vesi-etanoliseoksia käytetään hyvin yleisesti sekä teollisuudessa että kotitalouksissa. Siksi on ehkä hieman yllättävää, että niiden mikroskooppinen, molekyylitason rakenne ei ole vieläkään täydellisesti tunnettu. Tutkimuksissa on erityisesti tullut ilmi, että paikalliset molekyylitason rakenteet riippuisivat seossuhteesta. Tämän tutkimuksen tarkoituksena on pyrkiä saamaan lisätietoa vesi-etanoliseosten molekyylitason rakenteesta käyttäen kahta eri laskennallista menetelmää. Valitut menetelmät ovat klassinen molekyylidynamiikka (MD), jonka avulla voidaan tutkia molekyylien liikettä, sekä Compton-sironta, jossa sironnan vaikutusala riippuu elektronien liikemäärätiheydestä.

Tehdyissä MD-simulaatioissa alkoholipitoisuudet olivat välillä 0-100%. Vesimolekyylille ja etanolimolekyylille käytettiin niille yleisesti hyväksi osoittautuneita vuorovaikutusmalleja (TIP4P ja OPLS-AA). Lisäksi simulaatioissa käytettiin etanolin sidoksille kahta eri mallia, jäykkää ja joustavaa. Jäykässä mallissa molekyylien sisäiset sidospituudet ovat rajoitettuja, kun taas joustavassa mallissa sidospituudet määritetään harmonisella potentiaalilla. Kokeellisesti saatava suure, Compton-profiili, laskettiin kolmelle eri konsentraatiolle käyttäen sekä jäykkää että joustavaa etanolia.

Simulaatiot tuottivat hieman aliarvioidun tiheyden verrattuna kokeellisiin tuloksiin. Lisäksi sekä vesi- että etanolimolekyylien solmimien vetysidosten määrä poikkesi ideaalitapauksesta siten, että sidoksia muodostui enemmän veden kuin etanolin kanssa. Tutkimuksessa huomattiin myös, että seoksella oli järjestyneempi rakenne, kun etanolipitoisuus oli suurempi. Ero etanolimallien välillä oli merkityksetön MD-simulaatioissa.

Compton-sirontalaskut sen sijaan tuottivat selkeän eron eri etanolimallien välille. Jäykälle etanolille Compton-erotusprofiilit olivat kaikille etanolipitoisuuksille samat, mutta joustavalle etanolille ne erosivat toisistaan. Tämän perusteella sekoittumisprosessille on kaksi vaihtoehtoa. Vesi-etanoliseokset sekoittuvat samalla tavalla kaikilla konsentraatioilla (jäykän mallin tapaus) tai sekoittuminen muuttuu eri konsentraatioilla (joustavan mallin tapaus). Tämä tutkimus osoittaakin, että vuorovaikutusmallin valinnalla on merkittävä vaikutus MD-simulaatioista saatavan mikroskooppisen rakenteen muodostumiseen.

Kun Compton-profiilien laskennan tarkkuuteen vaikuttavia lähteitä arvoitiin, huomattiin, että laskuissa käytettävien MD-simulaatioista saatavien paikallisten rakenteiden otosmäärää on lisättävä, jotta tilastollinen virhe lopputuloksissa pienenisi. Tässä tutkimuksessa esitettyjä Compton-sirontatuloksia voidaankin pitää jossain määrin alustavina, mutta kuitenkin selkeästi suuntaa-antavina tuloksina vesi-etanoliseosten käyttäytymisestä, kun vuorovaikutusmallia muutetaan.

Tulevaisuudessa tutkimusta voidaan jatkaa lisäämällä näytteiden määrää ja vertaamalla tuloksia kokeellisiin mittauksiin, jolloin voidaan päättää, kumpi etanolimalleista kuvaa seosta paremmin. Siten on mahdollista saada merkittävää tietoa vesi-etanoliseosten mikroskooppisen tason rakenteesta. Sen lisäksi tietoa saadaan MD-simulaatioiden vuorovaikutusmallien valinnasta ja siitä, onko MD-simulaatiolla mahdollista luoda oikeita mikroskooppisia rakenteita kaksikomponenttinesteille.

Avainsanat — Nyckelord — Keywords
vesi-etanoliseos, molekyylidynamiikka, Compton-sironta

Säilytyspaikka — Förvaringsställe — Where deposited
Kumpulan tiedekirjasto

Muita tietoja — övriga uppgifter — Additional information

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Matemaattis-luonnontieteellinen tiedekunta	Fysiikan laitos

Tekijä — Författare — Author
Iina Juurinen

Työn nimi — Arbetets titel — Title
Water-ethanol mixtures by molecular dynamics and x-ray Compton scattering

Oppiaine — Läroämne — Subject
Teoreettinen fysiikka

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Pro gradu -tutkielma	Syyskuu 2009	51

Tiivistelmä — Referat — Abstract

Water-ethanol mixtures are commonly used in industry and house holds. However, quite surprisingly their molecular-level structure is still not completely understood. In particular, there is evidence that the local intermolecular geometries depend significantly on the concentration. The aim of this study was to gain information on the molecular-level structures of water-ethanol mixtures by two computational methods. The methods are classical molecular dynamics (MD), where the movement of molecules can be studied, and x-ray Compton scattering, in which the scattering cross section is sensitive to the electron momentum density.

Firstly, the water-ethanol mixtures were studied with MD simulations, with the mixture concentration ranging from 0 to 100%. For the simulations well-established force fields were used for the water and ethanol molecules (TIP4P and OPLS-AA, respectively). Moreover, two models were used for ethanol, rigid and non-rigid. In the rigid model the intramolecular bond lengths are fixed, whereas in the non-rigid model the lengths are determined by harmonic potentials. Secondly, mixtures with three different concentrations employing both ethanol models were studied by calculating the experimentally observable x-ray quantity, the Compton profile.

In the MD simulations a slight underestimation in the density was observed as compared to experiment. Furthermore, a positive excess of hydrogen bonding with water molecules and a negative one with ethanol was quantified. Also, the mixture was found more structured when the ethanol concentration was higher. Negligible differences in the results were found between the two ethanol models.

In contrast, in the Compton scattering results a notable difference between the ethanol models was observed. For the rigid model the Compton profiles were similar for all the concentrations, but for the non-rigid model they were distinct. This leads to two possibilities of how the mixing occurs. Either the mixing is similar in all concentrations (as suggested by the rigid model) or the mixing changes for different concentrations (as suggested by the non-rigid model). Either way, this study shows that the choice of the force field is essential in the microscopic structure formation in the MD simulations.

When the sources of uncertainty in the calculated Compton profiles were analyzed, it was found that more statistics needs to be collected to reduce the statistical uncertainty in the final results. The obtained Compton scattering results can be considered somewhat preliminary, but clearly indicative of the behaviour of the water-ethanol mixtures when the force field is modified.

The next step is to collect more statistics and compare the results with experimental data to decide which ethanol model describes the mixture better. This way, valuable information on the microscopic structure of water-ethanol mixtures can be found. In addition, information on the force fields in the MD simulations and on the ability of the MD simulations to reproduce the microscopic structure of binary liquids is obtained.

Avainsanat — Nyckelord — Keywords
water-ethanol mixtures, molecular dynamics, Compton scattering

Säilytyspaikka — Förvaringsställe — Where deposited
Kumpulan tiedekirjasto

Muita tietoja — övriga uppgifter — Additional information

Contents

1 Introduction

Water-ethanol mixtures are commonly used in industry and house holds. However, quite surprisingly their molecular-level structure is still not completely understood. Ethanol is easily soluble to water as its polar hydroxyl group can participate in the hydrogen bonded network and the hydrophobic ethyl group is relatively small. However, it has been shown experimentally that the water-alcohol mixing is incomplete at the molecular level, as there is negative excess entropy involved [1]. In this work classical molecular dynamics (MD) and x-ray Compton scattering (CS) are used to shed light on the microscopic structure of water-ethanol mixtures.

Water-ethanol mixtures have been used as drinking beverages for a long time. The earliest findings of alcohol production date back to Neolithic period, cir. 10 000 B.C. [2]. The first alcoholic beverages were made of fermented berries or honey [3]. Beer and wine have been everyday food products at least from 4 000 B.C. [4], providing fluid, calories and vitamins. Besides, because of ethanol's antiseptic properties, they were often safer to drink than water, which was often polluted and was dangerous or even fatal to drink. [5] Alcohol can be produced naturally up to concentration of only 14 vol-%, above which ethanol destroys the zymase enzyme, which is an enzyme from yeast that changes sugars into ethanol and carbon dioxide in the fermentation reaction [6]. In middle ages (circa 12th century) distillation was invented, which made it possible to produce alcohol products with significantly higher alcohol concentration [2, 5].

Alcohol has had a huge influence also as a medicine. Beer and wine were used for medicinal purposes as early as 2 000 B.C. in Sumer [7]. The main medicinal effects of alcohol were pain reduction and mood enhancement. Ethanol is also a good antiseptic, by being capable of destroying organisms by denaturing their proteins and dissolving their lipids. Ethanol is effective against most bacteria and fungi as well as many viruses. [8] After the invention of synthetic drugs, alcohol beverages have lost their place as a medicine [5], but still many medicines contain alcohol as a component (e.g. cough syrup), partly because of the antiseptic properties of ethanol and partly of the high ethanol solubility in many organic solvents [9].

Nowadays water-ethanol mixtures are widely used in industry. One of the new uses of water-ethanol mixtures is to replace fossil fuels [10]. Water-ethanol mixtures are good solvents and are used, for example, for perfumes, paints, deodorants and cleaning products.

Even though water-ethanol mixtures have been so widely used, still only little is known about the microscopic, atomic scale structure of the mixtures. The taste and smell of water-ethanol mixture depend strongly on the additional components and the mixing technique [11]. Also some properties of the mixtures show non-ideal behavior, meaning that the properties don't change according to the change in the concentration. Besides the negative excess entropy for water-ethanol mixtures, there is also, for example, large excess viscosity upon mixing [12]. This work focuses on the study of molecular-level mixing properties of water-ethanol solutions, in order to clarify the microscopic structures of the mixtures as obtained from a commonly used simulation methodology.

There has been several computational studies of water-ethanol mixtures [12–20]. The main focus of these studies has been structural or thermodynamical properties of the mixtures. For example, Wensink et al. [12] have reported excess heat, excess density, viscosity and mobility of water-ethanol mixtures. Noskov et al. [13] studied the hydration, dielectric and dynamic properties of water-ethanol mixtures, and discussed, for example, the role of hydrogen bonds. Hydrogen bonds in water-ethanol mixture have also been studied by van Erp et al. [19], and Zhang et al. [20]. Fidler et al. [16] studied the structure of water around ethanol via the radial distribution functions and coordination numbers.

Firstly, in this work classical molecular dynamics is used to study the molecular-level structure of water-ethanol mixtures. In classical molecular dynamics the movement of 10^x, with x usually less than 8, atoms is calculated numerically with classical (Newtonian or Lagrangian) dynamics. Molecular dynamics is often used to solve problems at the atomic scale which are hard to study experimentally or solve purely theoretically. [21] It can be used to compare results of model systems with experiments or compare theoretical predictions with results of model systems. Nowadays many new theories are first tested with simulations, before applying them in practise. [22]

The first molecular dynamics simulations were made in the late 1950's. The first simulations were accomplished for hard spheres [23, 24]. The first realistic molecular dynamics simulation was done in 1970's on liquid water [25]. Since then the molecular dynamics methodology has improved a lot, and because of the computing capacity nowadays it can, in some contexts, be even considered to be reliable enough to be a predictive tool [21].

In this work molecular dynamics simulations are used to study specifically the microscopic structure of water-ethanol mixtures which lay background to the macroscopic

properties of those liquids. Also, because of its ability to follow the coordinates of the atoms, molecular dynamics is used to obtain realistic local molecular geometries, which are then used as an input for the subsequent analysis.

Molecular level structure can be probed by various x-ray methods. Compton scattering, for example, is a well-established inelastic x-ray scattering technique [26]. In this technique, the experimentally observed scattering cross section can be interpreted using the electron momentum density of the system through the so-called Compton profile (CP). This way, detailed information on the local molecular and electronic structures can be obtained [27, 28].

Compton scattering was discovered already in 1920's by A. H. Compton, when he studied the scattering of γ-rays, but only in 1970's it has become a standard x-ray method [29]. Today Compton scattering experiments are done almost solely with synchrotron radiation, which produces high-flux and high-energy x-rays, making high-accuracy experiments possible [30].

The aim of this work is to study the microscopic structure of water-ethanol mixtures using, firstly, classical molecular dynamics and, secondly, computational Compton scattering methods. Combining these methods, additional information about the molecular dynamics method and on the accuracy of both methods will be received. In future these computational results could be compared with experimental data, which then will tell whether classical molecular dynamics can produce correct microscopical structures for the mixtures.

2 Classical molecular dynamics simulations

In classical molecular dynamics (MD) simulations the basic idea is to calculate how a system of particles evolves in time under the Newton's equations of motions. In MD one sets up a simulation cell with atoms or molecules. Every atom or molecule is assigned a force field, which is used to calculate the interactions between the other atoms and molecules. Also, temperature and pressure are controlled, scaling the velocities and positions of the atoms, respectively. With every time step the atoms are moved according to the forces of the current configuration. The movement of atoms is calculated using Newtonian or Lagrangian dynamics. The new coordinates, energies and other physical quantities may be stored. This way all the information can be saved over the whole simulation, and, for example, it is possible to see how the atom positions change during the simulation. In this section the principles of molecular dynamics method is discussed in more detail.

2.1 Simulation cell

The simulation cell (Fig. 1) can be generated in various ways. One of the most commonly used methods to create the initial configuration is to replicate and shift the already known monomer as many times as necessary. With binary systems the same procedure can be repeated with both of the monomers. Some of the MD software have ready-made tools to create the initial configuration [31]. The structures of the molecules can be obtained e.g. from the RCSB Protein Data Bank [32].

The initial atom velocities v_i are needed for certain algorithms, e.g. the leap-frog (discussed in Section 2.4) [33]. If the initial atom velocities are not known, they are created generally by using the Maxwell-Boltzmann distribution, with the given temperature T:

$$p(v_i) = \sqrt{\frac{m_i}{2\pi kT}} e^{-\frac{m_i v_i^2}{2kT}} \tag{1}$$

where k is the Boltzmann's constant and m_i the mass of the atom i. This can be done with random numbers. The resulting total kinetic energy, E_{kin}, can also be adjusted by scaling the velocities so that it corresponds to the required temperature,

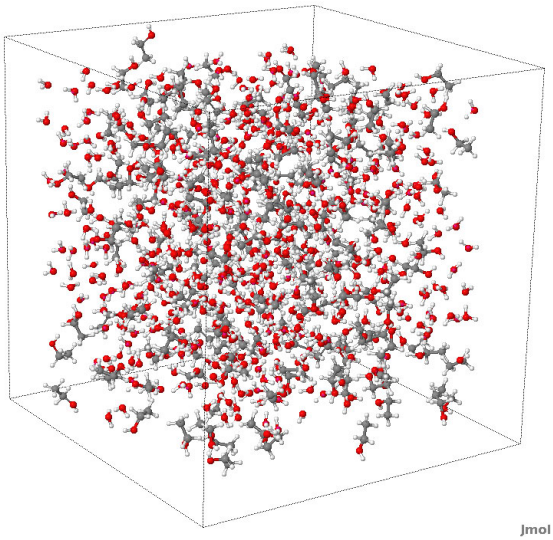

Figure 1: Simulation cell containing 500 water molecules and 192 ethanol molecules.

$$\frac{1}{2}N_{df}kT = E_{kin}, \tag{2}$$

where N_{df} is the number of the degrees of freedom in the system. [34]

Since the computing capacity is limited, the simulation cell is always of finite size. The state of the art supercomputers can calculate simulations with up to order of 10^{10} atoms [35], but for a commonly doable simulation the amount of atoms is at maximum of the order of 10^8 [21], which is still a small system from a macroscopic point of view. To avoid interactions at the simulation cell boundaries with the vacuum, periodic boundary conditions can be implemented. When using the periodic boundary conditions, the simulation cell is duplicated infinite times in every direction. In practice it means that when a molecule crosses the boundary, it will come back to the simulation cell from the other side. This is illustrated in Fig. 2. This also applies to interactions. There can be periodic boundaries in every direction, which corresponds to simulation of bulk material, or only in some directions. With two directions, the system corresponds to an infinite surface and with one direction to an infinite rod. [33]

The distance between two atoms is needed e.g. in force calculations. If the distances between every atom is calculated at every time step, the scaling of the algorithm is $O(N^2)$, where N is the amount of atoms [35]. However, in one time step, the atoms move only a small distance compared to the cut-off distance (i.e. the distance of how

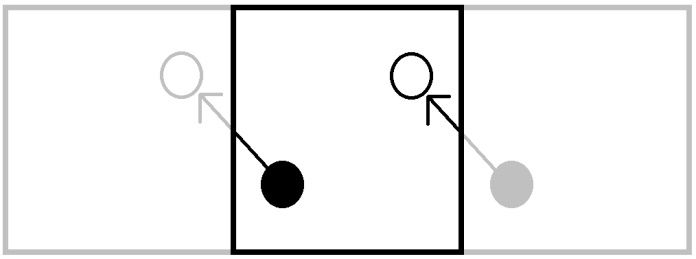

Figure 2: Movement through a periodic boundary.

far the interactions between the atoms are calculated, discussed in Section 2.4). So most of the neighbours stay approximately at the same positions during one time step, Δt, and there is no need for calculating them every step. The Verlet neighbour list [36] can be used to keep track of the neighbours. It is a list of the indices of the atoms j which are closer to atom i than a given distance R, which is larger than the cut-off distance r_{cut}. The list is updated every N_m time steps. The distance R and the interval N_m are chosen so that

$$R - r_{cut} > N_m \bar{v} \Delta t \tag{3}$$

where $\bar{v}$ is a typical atom velocity and Δt the time step. Using this method a factor of N_m can be saved in time but still the whole algorithm is $O(N^2)$. [33]

The algorithm can be made to $O(N)$ by using a cell method, in which the simulation cell is divided into $M \times M \times M$ subcells. The size of one subcell l is chosen so that

$$l = \frac{L}{M} > R \tag{4}$$

where L is the size of the simulation cell and R is the Verlet list distance. To search the neighbours of atom i, only the subcell where the atom is and the neighbouring subcells has to be searched through. When the average number of atoms in a subcell is $N_c = N/M^3$, only $27NN_c$ atom pairs have to be gone through instead of $N(N-1)$. This way the algorithm scales as $O(N)$. [33]

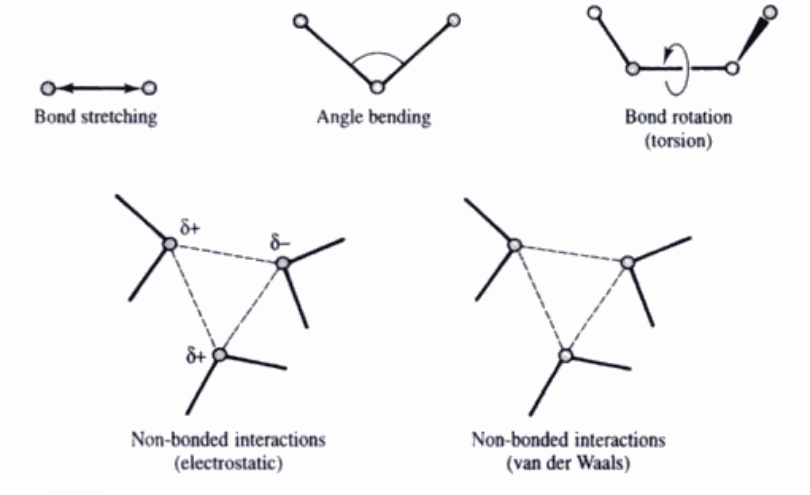

Figure 3: Key contributions to the force fields. [37]

2.2 Force fields

The interactions between the atoms can be divided into two groups, bonded and nonbonded interactions. Bonded interactions are interactions of atoms which are bonded to each other i.e. are in the same molecule. Nonbonded interactions are interactions between atoms in different molecules, but these interactions are generally also applied to bonded atoms, especially in long molecules. The interactions are presented in Fig. 3. The total potential generally consists of five parts

$$V = V_{bonds} + V_{angles} + V_{torsional} + V_{electrostatic} + V_{Lennard-Jones}, \tag{5}$$

where the first term is the bond stretching potential, second the bond angle potential and third the torsional potential, i.e., the rotation of the molecule bonds. These three are the bonded interactions. In equation 5 the remaining two terms are nonbonded interactions. The first is the electrostatic potential and the last one the Lennard-Jones potential. [37]

One of the simplest ways to describe the bond stretching potential is the harmonic potential

$$V_{bonds} = \frac{1}{2} \sum_{bonds} k_i (l_i - l_{i,0})^2, \tag{6}$$

where k_i is a force constant and the bond length, l_i, oscillates around the reference length $l_{i,0}$. [37]

The bond angle potential is generally treated similarly to bond stretching potential as a harmonic potential,

$$V_{bond\ angles} = \frac{1}{2} \sum_{bond\ angles} k_i'(\theta_i - \theta_{i,0})^2, \tag{7}$$

where k_i' is a force constant and the bond angle, θ_i, oscillates similarly around the reference angle $\theta_{i,0}$. [37]

The torsional potential has a substantial effect in the bond energy especially in long molecules [38]. It is generally described as a cosine expansion

$$V_{torsion} = \frac{1}{2} \sum_{torsion} V_n(1 + \cos(n\varphi - \delta)), \tag{8}$$

where φ is the torsion angle and V_n a constant. In equation 8 n is the multiplicity, i.e., it describes how many minimum points there are in the function as the bond is rotated through 360°. [37]

Different electronegativity in atoms causes uneven charge distribution in the molecules. The most common way to describe this interaction is the Coulomb potential

$$V_{Coulomb} = \sum_{i,j} \frac{1}{4\pi} \frac{q_i q_j}{\varepsilon_0 r_{ij}}, \tag{9}$$

where q_i and q_j are point charges, ε_0 the permittivity of free space and r_{ij} the distance between the point charges.

All nonbonded interactions cannot be explained via electronegativity, so something else is needed. These interactions are called van der Waals interactions and are commonly described using the Lennard-Jones potential

$$V_{LJ} = \frac{1}{2} \sum_{i,j} 4\epsilon_{ij}((\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}). \tag{10}$$

When binary systems are studied, where there are atoms of different types i and j, the two atom type Lennard-Jones parameters, ϵ_{ij} and σ_{ij}, can be calculated following Lorentz and Bertelot, where [31, 39]

$$\epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{\frac{1}{2}}, \tag{11}$$

and

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}), \tag{12}$$

where ϵ_{ii}, ϵ_{jj},σ_{ii} and σ_{jj} are the Lennard-Jones parameters for interactions between the same atom types. These interactions are discussed in more detail in the author's bachelor thesis [40].

Basically all the interactions range to infinity. However, the bonded interactions are restricted to the closests neighbours, so only the non-bonded interactions can range to infinity. Because of the limited computing capacity and the finite size simulation cell, the interactions must be in practice truncated. This is generally solved by using cut-offs with the interactions. The interactions are calculated only up to a certain distance and after this cut-off distance the potential of the interaction goes to zero. The size of the simulation cell restricts the cut-off distance, because due to the periodic boundaries a molecule can interact with itself if the cut-off distance is too large. Basically the cut-off distance has to be smaller than half of the shortest unit vector of the simulation cell. [33]

In classical molecular dynamics the forces are calculated in the Newtonian way and the force is a negative gradient of the total potential

$$-\nabla \mathbf{V}_i = \mathbf{F}_i = m_i \mathbf{a}_i = m_i \frac{d^2 \mathbf{r}_i}{dt^2}, \tag{13}$$

where $\mathbf{F}_i$ is the force acting on atom i with mass m_i [33].

2.3 Energy minimization

Before the actual molecular dynamics run, the energy of the initial configuration has to be minimized, otherwise there can be huge effects from e.g. two atoms which are too close to each other. One of the ways to minimize energy is the steepest descent -method, where first the forces and the potential energy are calculated, and after that the new positions for the atoms ($\mathbf{r}_i$) are calculated by

$$\mathbf{r}_i^{n+1} = \mathbf{r}_i^n + \frac{\mathbf{F}_i^n}{max(|\mathbf{F}_i^n|)} h^n, \tag{14}$$

where h^n is the maximum displacement from the previous position, $\mathbf{F}_i^n$ is the force and $max(|\mathbf{F}_i^n|)$ is the largest of absolute values of the force components. Above, n denotes the steepest descent step. The initial maximum displacement h^0 must be given. Then the forces and the potential energies are calculated for the new positions and if the potential is smaller than in the previous step ($V^{n+1} < V^n$) the new positions are accepted and $h^{n+1} = 1.2 \cdot h^n$ or if the potential is larger than in the previous step ($V^{n+1} \geq V^n$) the new positions are rejected and $h^{n+1} = 0.2 \cdot h^n$. The method is repeated until a specified number of iterations or until the forces are smaller than a specified value. [31, 41]

Another way to minimize energy is the conjugate gradients method, where instead of moving the atom to the direction of the force, it is moved with two vectors, one perpendicular to the force, and another one perpendicular to that. The atoms are moved by [42]

$$\mathbf{r}_i^{n+1} = \mathbf{r}_i^n + \kappa \cdot \mathbf{l}_i^{n+1}, \tag{15}$$

where κ is constant and

$$\mathbf{l}_i^{n+1} = \mathbf{F}(\mathbf{r}_i^{n+1}) + \gamma_i^n \mathbf{l}_i^n, \tag{16}$$

and γ is updated as

$$\gamma_i^n = \frac{\mathbf{F}(\mathbf{r}_i^{n+1}) \cdot \mathbf{F}(\mathbf{r}_i^{n+1})}{\mathbf{F}(\mathbf{r}_i^n) \cdot \mathbf{F}(\mathbf{r}_i^n)}. \tag{17}$$

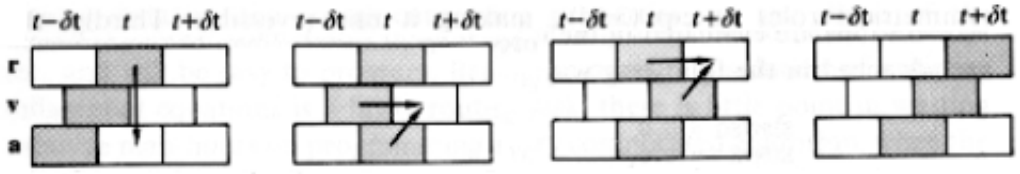

Figure 4: Leap-frog algorithm, where the system evolves over one time step $\Delta t \equiv \delta t$. Firstly from the atom positions at time t one calculates the accelerations. Then using the velocity from the previous half-time step (at time $t - \Delta t/2$) and acceleration at time t the new velocity at time $t + \Delta t/2$ is calculated. Then using that velocity and position from time t, the new position at time $t + \Delta t$ is calculated. [33]

2.4 Equations of motion

The main goal in the MD simulation is to solve the equations of motion. There are several numerical algorithms to do it. One commonly used is the leap-frog algorithm, where the positions and the accelerations are calculated at times t and the velocities at times $t - \frac{\Delta t}{2}$, where Δt is the time step of the simulation [33]

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t + \frac{1}{2}\Delta t), \tag{18}$$

$$\mathbf{v}_i(t + \frac{1}{2}\Delta t) = \mathbf{v}_i(t - \frac{1}{2}\Delta t) + \Delta t \mathbf{a}_i(t). \tag{19}$$

The accelerations are obtained from the forces of the current time according to equation 13.

The way to calculate the accelerations, velocities and positions is illustrated in Fig. 4. To calculate e.g. the energy at time t, the velocities at that time are needed. The velocities at time t are averages of the previous and the following half time steps [33]

$$\mathbf{v}_i(t) = \frac{1}{2}(\mathbf{v}_i(t + \frac{1}{2}\Delta t) + \mathbf{v}_i(t - \frac{1}{2}\Delta t)). \tag{20}$$

2.5 Temperature and pressure control

In order to make the simulation correspond to a realistic case, one has to characterize the macroscopic equilibrium state by the state parameters like temperature. Often the MD simulations are done in the isothermal-isobaric ensemble, also called the NTP-ensemble [21], where the number of atoms, the temperature and the pressure are kept constant. There are many methods to keep the temperature constant, one of which is the Berendsen temperature scaling.

The Berendsen temperature scaling method is essentially a direct scaling but it is softened with a time constant [43]. When coupled to a heat bath, the temperature is corrected according to

$$\frac{dT}{dt} = \frac{1}{\tau}(T_0 - T), \tag{21}$$

where T_0 is the desired temperature and τ is a time constant. So the temperature deviation decays exponentially with τ. [31, 43]

To make the heat flow into or out of the system, the particle velocities are scaled. At every time step the particle velocities are scaled by a time dependent factor

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left\{\frac{T_0}{T(t - \frac{\Delta t}{2})}\right\}\right]^{\frac{1}{2}}, \tag{22}$$

where Δt is the time step and the parameter τ_T is proportional to the time constant τ,

$$\tau = \frac{2C_V\tau_T}{N_{df}k}, \tag{23}$$

where C_V is the total heat capacity of the system. [31, 43]

Similarly to temperature, also pressure can be scaled. Pressure is controlled by scaling the coordinates and vectors defining the simulation cell at every time step. One of the commonly used methods to control the pressure is the Berendsen pressure coupling.

The Berendsen method scales the coordinates and the simulation cell vectors every time step with a matrix μ. The pressure, $\mathbf{P}$, is scaled similarly to temperature (equation 21) towards the given reference pressure $\mathbf{P}_0$

$$\frac{d\mathbf{P}}{dt} = \frac{1}{\tau_p}(\mathbf{P}_0 - \mathbf{P}). \tag{24}$$

The scaling matrix is given by the elements

$$\mu_{lm} = \delta_{lm} - \frac{\Delta t}{3\tau_p}\beta_{lm}(P_{0lm} - P_{lm}(t)), \tag{25}$$

where β is the isothermal compressibility of the system, δ_{lm} the Kronecker delta and l and m represent the cartesian directions x, y and z. [31, 44]

2.6 Analysis tools

There are lots of different quantities that can be studied from the MD simulation. In this section those relevant to this work are described.

The radial distribution function (RDF) (also called the pair correlation function) describes how the density of the surrounding matter varies as a function of the distance from a particular point [45]. Experimentally it can be obtained e.g. with x-ray or neutron diffraction [46]. In a binary system the radial distribution between two atom species A and B can be defined as

$$g_{AB}(r) = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{local}} = \frac{1}{\langle \rho_B \rangle_{local}} \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta(r_{ij} - r)}{4\pi r^2}, \tag{26}$$

where $\langle \rho_B(r) \rangle$ is the atom density of atom species B at a distance r around atoms A, $\langle \rho_B \rangle_{local}$ is the atom density of atoms B averaged over all spheres around atoms A within the maximum distance r_{max} and N_A and N_B are the number of atoms A and B, respectively. The maximum distance r_{max} is usually half of the simulation cell length. [31]

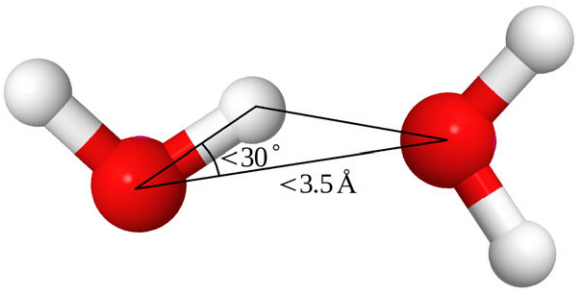

Figure 5: Definition of hydrogen bond used in this work. The indicated distance must be less than 3.5 Å and the angle less than 30 degrees. The water molecule to the left is the donor and the molecule to the right the acceptor.

The coordination number describes how many closest neighbours an atom has. It can be estimated from the RDF by integrating over the first peak to the first minimum at R_{min}. The coordination number of the atom species B around those of A is

$$N_{AB}^{coord} = 4\pi\rho_B \int_0^{R_{min}} R^2 g_{AB}(R) dR, \tag{27}$$

where ρ_B is the mean density of atoms B in the system. [47]

There are also methods for calculating the average number of hydrogen bonds, N_{HB}. There is no universally agreed definition of what is a hydrogen bond. In this work the hydrogen bond is defined in the following way, which is often used: the distance between the oxygens is less than 3.5 Å and the angle acceptor-donor-hydrogen is less than 30°. See Fig. 5 for the definition of the distances, angles and the definition of an acceptor and a donor molecule.

From the simulations the hydrogen bonds can be calculated so that the number of hydrogen bonds per time step is calculated, and then average of those over the simulation time is the total average number of hydrogen bonds in the simulation cell. From the total number of hydrogen bonds the number of hydrogen bonds per one molecule can be calculated by dividing the total number by the amount of the molecules of the desired molecule type in the simulation cell. This number has to be multiplied by the factor of two if the bond is calculated between same type of molecules.

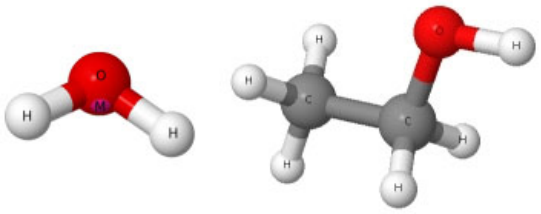

Figure 6: Left: TIP4P water molecule with the virtual charge M. Right: OPLS-AA ethanol molecule.

2.7 Molecular dynamics setup

In this work for water the TIP4P (Transferable Intermolecular Potential, 4 Points) [48] force field was used. In this model the water molecule is described with four point charges; in addition to the atom charges (oxygen and hydrogens) there is a virtual charge M, which shifts the total charge (Fig. 6). For ethanol the OPLS-AA (Optimized Potential for Liquid Simulation - All Atom) [49] force field was used, where all the atoms are treated individually (Fig. 6). Two approaches were studied for ethanol, denoted as 'rigid' and 'non-rigid'. For the rigid ethanol the LINCS-algorithm [50] was used to maintain the geometry of the molecule. For the non-rigid ethanol no constraints were used, except those defined explicitly in the topology, i.e. bonds were represented by a harmonic potential. More information about the water and the ethanol models can be found in the author's bachelor thesis [40].

For all the MD simulations the Gromacs software [51] was used. All the simulations were for 2 ns and prior to that there was an equilibration simulation for 2 ns. The time step was 1 fs for non-rigid ethanol and 2 fs for rigid ethanol both with the leap-frog integrator. The cut-offs for the potentials and the neighbourlists were 1.1 nm, except for the mixtures with 0.5% of ethanol concentration this was chosen to be 0.9 nm (the simulation cells were not large enough for longer cut-off lengths). To keep the temperature and pressure constant the Berendsen algorithm was used. The reference temperature was 300 K and the reference pressure 1 bar. The coupling constant was 0.1 ps for the temperature and 0.5 ps for the pressure. The neighbourlists were updated every 5th time step and the energies and coordinates were stored every 100 fs.

The ethanol concentration in the simulations ranged from 0% to 100% (mass percents). Table 1 lists the used concentrations (both mass and molar concentrations). Also in Table 1 are listed the used ethanol model, the amount of ethanol and water molecules, the cut-off-lengths, and the densities obtained from the simulations. These results are discussed in Section 4.1.

Table 1: Setup used in the MD simulations. The ethanol model, amounts of ethanol and water molecules in the simulation cell, ethanol molar and mass concentration, cut-off and the obtained density and the standard deviation from the simulation are given.

EtOH-model	**#EtOH**	**#H_2O**	**EtOH mol-%**	**EtOH m-%**	**Cut-off (nm)**	**Density (g/l)**
-	0	1016	0	0	1,1	(1003,2±5,4)
Non-rigid	1	211	0.5	1.2	0.9	(1001.6±11.4)
Non-rigid	36	782	4.4	10.53	1.1	(979.2±5.3)
Non-rigid	75	745	9.1	20.47	1.1	(958.2±5.1)
Non-rigid	120	620	16.2	33.11	1.1	(931.7±5.5)
Non-rigid	160	519	23.6	44.08	1.1	(909.9±5.6)
Non-rigid	192	538	26.3	47.71	1.1	(903.4±5.7)
Non-rigid	240	400	37.5	60.54	1.1	(876.8±5.5)
Non-rigid	280	300	48.3	70.47	1.1	(856.0±5.8)
Non-rigid	315	200	61.2	80.11	1.1	(835.5±5.8)
Non-rigid	360	100	78.3	90.20	1.1	(812.3±5.7)
Non-rigid	360	20	94.7	97.87	1.1	(792.5±6.0)
Non-rigid	392	0	100	100	1.1	(787.3±5.5)
Rigid	1	211	0.5	1.2	0.9	(1003.1±11.5)
Rigid	36	842	4.1	9.86	1.1	(983.6±5.4)
Rigid	75	729	9.3	20.83	1.1	(962.7±5.3)
Rigid	118	597	16.5	34.26	1.1	(938.0±5.3)
Rigid	158	600	20.8	40.24	1.1	(925.4±5.3)
Rigid	192	500	27.7	49.54	1.1	(906.2±5.3)
Rigid	239	400	37.4	60.44	1.1	(884.2±5.1)
Rigid	280	300	48.3	70.47	1.1	(863.7±5.4)
Rigid	315	200	61.2	80.11	1.1	(843.3±5.1)
Rigid	360	100	78.3	90.20	1.1	(819.5±5.2)
Rigid	360	20	94.7	97.87	1.1	(799.7±5.9)
Rigid	392	0	100	100	1.1	(794.1±5.5)

3 X-ray Compton scattering

X-ray Compton scattering (CS) is inelastic scattering far from the resonances of the system when there is a large energy and momentum transfer from the photon to the target electron [52]. X-ray Compton scattering can be used to study the structure of materials at the atomic scale, since it probes the electronic orbitals of the system in the momentum space. [29] In this section the theoretical model for calculating Compton scattering cross sections is presented.

3.1 Theoretical model

Inelastic x-ray scattering is a photon-in photon-out process described schematically in Fig. 7 [26]. A photon of energy $\hbar\omega_1$, which has the wave vector $\mathbf{k}_1$ and the polarization unit vector $\hat{\varepsilon}_1$, collides with a target which is characterized by a state vector $|I\rangle$ and initial energy E_I. The outgoing photon has an energy $\hbar\omega_2$, wave vector $\mathbf{k}_2$ and polarization unit vector $\hat{\varepsilon}_2$. This leaves the target in the final state $|F\rangle$ with energy E_F. An energy $\hbar(\omega_1 - \omega_2) = \hbar\omega$ and a momentum $\hbar\mathbf{k} = \hbar(\mathbf{k}_1 - \mathbf{k}_2)$ are transferred to the target. Energy conservation requires that

$$\hbar\omega = E_F - E_I. \tag{28}$$

The amount of transferred momentum is connected with the scattering angle ϕ by

$$k = \frac{1}{c}(\omega_1^2 + \omega_2^2 - 2\omega_1\omega_2\cos\phi)^{\frac{1}{2}}, \tag{29}$$

where c is the speed of light. If the transferred energy is much smaller than the initial photon energy, $\omega \ll \omega_1$, equation 29 reduces to

$$k \approx 2k_1 \sin\left(\frac{\phi}{2}\right). \tag{30}$$

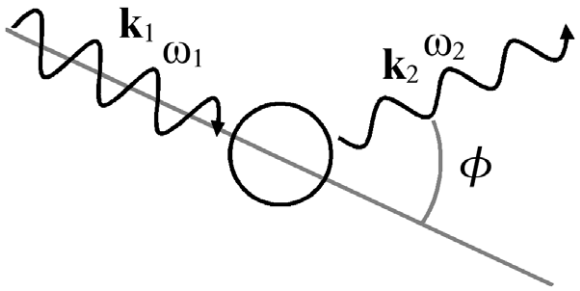

Figure 7: Schematic picture of Compton scattering.

In typical Compton scattering experiments the incoming beam is well-collimated and consists of monochromatic photons. A certain solid angle $d\Omega$ is selected of the scattered beam (this fixes $\mathbf{k}$ according to equation 30). So the double differential scattering cross-section $\frac{d^2\sigma}{d\Omega d\omega_2}$ is measured as a function of $\mathbf{k}$ and ω. In the non-relativistic limit the double differential scattering cross-section for inelastic x-ray scattering consists of two factors,

$$\frac{d^2\sigma}{d\Omega d\omega_2} = \left(\frac{d\sigma}{d\Omega}\right)_0 S(\mathbf{k},\omega), \tag{31}$$

where $S(\mathbf{k},\omega)$ is the dynamic structure factor, which reflects the properties of the target without the perturbing probe, and $\left(\frac{d\sigma}{d\Omega}\right)_0$ is the Thomson scattering cross-section, which describes the coupling of the electron to the electromagnetic field. The separation in equation 31 can be made provided that the coupling to the electromagnetic field is weak enough to be treated in the lowest-order Born approximation (first order perturbation theory) and the resonance phenomena can be neglected. [26] The Thomson scattering cross-section can be written in the following way [26, 53, 54]

$$\left(\frac{d\sigma}{d\Omega}\right)_0 = r_0 \left(\frac{\omega_2}{\omega_1}\right) (\hat{\varepsilon}_1 \cdot \hat{\varepsilon}_2)^2, \tag{32}$$

where r_0 is the classical electron radius.

For many-electron systems the Fermi's golden rule leads to the formula

$$S(\mathbf{k},\omega) = \sum_F \left| \langle F| \sum_{l=1}^{Z} e^{i\mathbf{k}\cdot\mathbf{r}_1} |I\rangle \right|^2 \delta(E_F - E_I - \hbar\omega), \tag{33}$$

where $\mathbf{r}_l$ is the position of the lth electron [29]. For the delta function an integral representation can be used [26]

$$\delta(E_F - E_I - \hbar\omega) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} e^{-it/\hbar(E_F - E_I - \hbar\omega)}dt. \tag{34}$$

With this the following expression for the dynamic structure factor can be obtained [26]

$$S(\mathbf{k},\omega) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dt e^{-i\omega t}\sum_F \langle I|\sum_l e^{-i\mathbf{k}\cdot\mathbf{r}_l}|F\rangle \times \langle F|e^{iHt/\hbar}\sum_l e^{i\mathbf{k}\cdot\mathbf{r}_l}e^{-iHt/\hbar}|I\rangle. \tag{35}$$

Above, $H|x\rangle = E_x|x\rangle$ is used, where H is the Hamiltonian and the state $|x\rangle$ is either $|I\rangle$ or $|F\rangle$.The Hamiltonian operator H in equation 35 can be divided into the kinetic energy term H_0 and the potential energy term V. The exponent function in equation 35 can be expanded in the following way

$$e^{iHt/\hbar} = e^{iH_0t/\hbar}e^{iVt/\hbar}e^{-[H_0,V]t^2/(2\hbar^2)}... \tag{36}$$

The high order terms contain multiple commutators and are of higher order in t. In the impulse approximation it is assumed that whenever

$$\hbar\omega \gg (\langle[H_0,V]\rangle)^{\frac{1}{2}}, \tag{37}$$

one can approximate

$$e^{-[H_0,V]t^2/(2\hbar^2)} \cong 1, \tag{38}$$

since significant contribution to the time integral in equation 35 occur only for $t \leq 1/\omega$. [26]

For clarity, in the following we consider a one-electron atom and drop the sum $\sum_l$. Since the potential V commutes with $\mathbf{r}$, by using equation 38, the dynamic structure factor will be

$$S(\mathbf{k},\omega) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dt e^{-i\omega t}\langle I|e^{-i\mathbf{k}\cdot\mathbf{r}}e^{iH_0 t/\hbar}e^{\mathbf{k}\cdot\mathbf{r}}e^{-iH_0 t/\hbar}|I\rangle. \tag{39}$$

When the complete set of eigenfunctions $|\mathbf{p}_f\rangle$ of the kinetic energy part of the Hamiltonian with

$$e^{iH_0 t/\hbar}|\mathbf{p}_f\rangle = e^{iE(\mathbf{p}_f)t/\hbar}|\mathbf{p}_f\rangle, \tag{40}$$

where

$$E(\mathbf{p}_f) = \mathbf{p}_f^2/2m \tag{41}$$

is applied to equation 39, it will transform into

$$S(\mathbf{k},\omega) = \sum_{\mathbf{p}_f}\left|\langle I|e^{-i\mathbf{k}\cdot\mathbf{r}}|\mathbf{p}_f\rangle\right|^2 \delta[E(\mathbf{p}_f) - E(\mathbf{p}_f - \hbar\mathbf{k}) - \hbar\omega]. \tag{42}$$

If $\mathbf{p}$ is defined as

$$\mathbf{p} \equiv \mathbf{p}_f - \hbar\mathbf{k} \tag{43}$$

and the sum over the final states $|\mathbf{p}_f\rangle$ is switched to a $\mathbf{p}$ integration, the dynamic structure factor becomes

$$S(\mathbf{k},\omega) = \left(\frac{1}{2\pi\hbar}\right)^3 \int |\langle I|p\rangle|^2 \delta(\hbar^2k^2/2m + \hbar\mathbf{p}\cdot\mathbf{k}/m - \hbar\omega)d\mathbf{p}. \tag{44}$$

When $\varphi(\mathbf{r})$ is the single particle position space wave function representing the initial state $|I\rangle$, then its Fourier transform is $\chi(\mathbf{p})$ and we have

$$\left(\frac{1}{2\pi\hbar}\right)^3 |\langle I|\mathbf{p}\rangle|^2 = |\chi(\mathbf{p})|^2. \tag{45}$$

With this, the dynamic structure factor becomes

$$S(\mathbf{k},\omega) = \int |\chi(\mathbf{p})|^2 \delta(\hbar^2 k^2/2m + \hbar\mathbf{p}\cdot\mathbf{k}/m - \hbar\omega) d\mathbf{p}. \tag{46}$$

The same result (equation 46) can be obtained also in another way, which is here briefly reviewed [29]. In the Van Hove representation [55] the dynamic structure factor of equation 35 is described as

$$S(\mathbf{k},\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\omega t} \langle I|\rho_{\mathbf{k}}(t)\rho_{-\mathbf{k}}(0)|I\rangle dt. \tag{47}$$

The density operator is defined by

$$\rho_{\mathbf{k}} \equiv \sum_l e^{i\mathbf{k}\cdot\mathbf{r_l}}, \tag{48}$$

$$\rho_{\mathbf{k}}(t) = e^{iHt/\hbar}\rho_{\mathbf{k}}e^{-iHt/\hbar}. \tag{49}$$

Assuming equations 37 and 38 and noticing that the terms in $e^{iVt/\hbar}$ cancel out, the time dependent density operator can be estimated by approximation of the Hamiltonian. The density operator is then

$$\rho_{\mathbf{k}}(t) \approx e^{iH_0t/\hbar}\rho_{\mathbf{k}}e^{-iH_0t/\hbar}, \tag{50}$$

where H_0 is the free electron Hamiltonian. [29]

With annihilation $a_{\mathbf{p}}$ and creation $a^+_{\mathbf{p}}$ operators for free electrons with momentum $\mathbf{p}$, the density operator becomes

$$\sum_l e^{i\mathbf{k}\cdot\mathbf{r_l}} = \sum_p a^+_{\mathbf{p}} a_{\mathbf{p}+\hbar\mathbf{k}}. \tag{51}$$

The time dependence of the annihilation and the creation operators can be approximated similarly to equation 50. Using these approximations the dynamic structure factor becomes

$$S(\mathbf{k},\omega) = \sum_{\mathbf{p}} \sum_{\mathbf{p}'} \delta(\hbar^2 k^2/2m + \hbar\mathbf{p}\cdot\mathbf{k}/m - \hbar\omega)\langle a^+_{\mathbf{p}} a_{\mathbf{p}+\hbar\mathbf{k}} a^+_{\mathbf{p}'+\hbar\mathbf{k}} a_{\mathbf{p}'}\rangle. \tag{52}$$

Both $\mathbf{p}$ and $\mathbf{p}'$ are of order $\mathbf{p}_c$, which is a characteristic ground state momentum, and the momentum transfer $\hbar\mathbf{k}$ is large compared to the ground state momentum $\hbar\mathbf{k} \gg \mathbf{p}_c$. The last part of equation 52 then becomes

$$\sum_{\mathbf{p}'} \langle a^+_{\mathbf{p}} a_{\mathbf{p}+\hbar\mathbf{k}} a^+_{\mathbf{p}'+\hbar\mathbf{k}} a_{\mathbf{p}'}\rangle \approx \langle a^+_{\mathbf{p}} a_{\mathbf{p}}\rangle, \tag{53}$$

which is just the momentum density $\langle a^+_{\mathbf{p}} a_{\mathbf{p}}\rangle = n(\mathbf{p})$.

So now the dynamic structure factor is

$$S(\mathbf{k},\omega) = \left(\frac{1}{2\pi\hbar}\right)^3 \int n(\mathbf{p})\delta(\hbar^2 k^2/2m + \hbar\mathbf{p}\cdot\mathbf{k}/m - \hbar\omega)d\mathbf{p}, \tag{54}$$

which is the same as equation 46 if $n(\mathbf{p}) = |\chi(\mathbf{p})|^2$. [29]

Let us now proceed to calculate $n(\mathbf{p})$ for an electron system in an ionic potential. The basic idea in the density functional theory is to describe the electron system by its density instead of using the many-body wave function. It is generally used to describe the system in the ground state. For N electrons the benefit of using the density functional theory is that the basic variable depends only on three degrees of freedom instead of $3N$. The ground state density in the Kohn-Sham density functional theory is calculated as

$$n(\mathbf{r}) = \sum |\varphi_j(\mathbf{r})|^2, \tag{55}$$

where $\{\varphi_j(\mathbf{r})\}$ is the orthonormal set of Kohn-Sham single particle wave functions in the ground state and the sum is over occupied states. The same formulation also

follows from the Hartree-Fock approximation. [56] The momentum density is the sum of the Fourier-transformed wave functions [27]:

$$n(\mathbf{p}) = \sum_j (2\pi\hbar)^{-3} \left| \int e^{-i\mathbf{p}\cdot\mathbf{r}/\hbar} \varphi_j(\mathbf{r}) d\mathbf{r} \right|^2 = \sum_j |\chi_j(\mathbf{p})|^2. \tag{56}$$

The method of calculation described in this section is called the impulse approximation, where the scattering takes place so quickly that other electrons cannot participate in the process and the incoming photon only collides with a moving single electron [57].

Integrating the equation 54 of the dynamic structure factor is easy and if $\mathbf{k}$ is chosen to be in z-direction, one can define the Compton profile (CP)

$$J(p_z) = \int_{p_x} \int_{p_y} n(\mathbf{p}) dp_x dp_y. \tag{57}$$

If the system is isotropic, $n(\mathbf{p}) \equiv n(p)$, a spherically averaged Compton profile is obtained as

$$J(q) = \frac{1}{2} \int_{|q|}^{\infty} \int_0^{2\pi} \int_0^{\pi} n(p) \sin\theta d\theta d\phi p dp = \frac{1}{2} \int_{|q|}^{\infty} 4\pi p n(p) dp. \tag{58}$$

The Compton profile is generally normalized as

$$\int_{-\infty}^{\infty} J(q) dq = Z, \tag{59}$$

where Z is the number of electrons in one molecule. [26, 57]

So finally we get for the cross-section of Compton scattering [29]

$$\frac{d^2\sigma}{d\Omega d\omega_2} = \left(\frac{d\sigma}{d\Omega} \right)_0 \frac{m}{|\hbar k|} J(q). \tag{60}$$

Relativistic kinematics affects the connection between the scalar variable q and the experimental parameters, the photon energies $\hbar\omega_1$ and $\hbar\omega_2$, and the scattering angle ϕ. A good approximation of the connection is

$$q \simeq \frac{\hbar|\mathbf{k}|}{2} - \frac{\hbar(\omega_1 - \omega_2)}{c}\sqrt{\frac{1}{4} - \frac{m^2c^4}{2\hbar^2\omega_1\omega_2(1 - cos\phi)}}, \tag{61}$$

where c is the speed of light. [53]

At this level of approximation these equations can be used to interpret experimental data [29]. In the cross-section (equation 60) everything except the Compton profile depends only on the experimental set-up [27].

3.2 Calculation of Compton profiles from MD structures

Since one of the benefits of using MD simulation is that the coordinates of the atoms are recorded over the simulations, it is convenient to use the local geometries from the MD simulations as an input for the Compton profile calculations.

Compton profiles are calculated using quantum mechanical description for the electronic structure, but the calculations are much heavier than for classical molecular dynamics. Because of the limited computing time, in the Compton profile calculations only small clusters extracted from the simulation cell can be calculated. For pure liquids, clusters with 20 water molecules (for pure water) or 8 ethanol molecules (for pure ethanol) were randomly selected at 5-10 different time steps. For mixtures the clusters are sampled so that all molecules within the radius of 5.5 Å are included at 5-10 different time steps. Altogether, about 100 clusters have been randomly selected for one mixture. This means that the amount of the molecules in one mixture cluster varies, and also the geometries vary. This way an overall sampling of the molecular configurations can be obtained. A representative cluster from the mixture is shown in Fig. 8.

The Compton profile calculation is done with the StoBe-deMon software [58], which employs localized molecular orbitals, density-functional theory and Kohn-Sham single particle states.

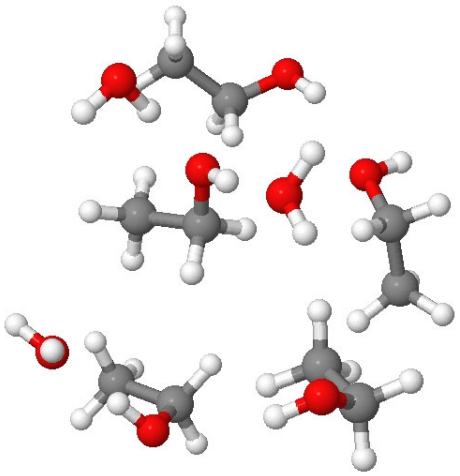

Figure 8: Representative cluster used in the Compton profile calculation.

In this work the main focus on Compton profiles is in the difference profiles

$$\Delta J(q) = J(q) - J'(q), \tag{62}$$

where $J(q)$ is the Compton profile of the mixture and $J'(q)$ is the weighted sum of the Compton profiles of pure water and ethanol. This is the quantity that can be obtained from experiments with high statistical precision. For example, recently a precision of 0.01%-units has been obtained for the ratio $\frac{\Delta J(q)}{J(0)}$ [28].

3.3 Bootstrap method

The bootstrap method is suitable to roughly estimate the statistical errors in a set of variables. In this work the bootstrap method is used to approximate errors in the Compton profiles. The procedure is the following:

Let there be a set of N spectra, $\{S_n(\omega)\}_{n=1}^{N}$. An average spectrum is calculated as

$$\langle S(\omega) \rangle = \frac{1}{N} \sum_{n=1}^{N} S_n(\omega). \tag{63}$$

Then M bootstrap spectra are calculated from the N spectra (creating a set of bootstrap spectra $\{B_n(\omega)\}_{n=1}^{M}$). The bootstrap spectra are averages of groups of randomly

selected original spectra, so that N spectra from the group $\{S_n(\omega)\}_{n=1}^{N}$ are selected randomly, discarding that one spectrum can be selected multiple times. Then from that group of spectra the average spectrum is the bootstrap spectrum. The M bootstrap spectra can be statistically analyzed or drawn in the same figure. The standard deviation from the average spectrum can be then used as an error estimate. [59]

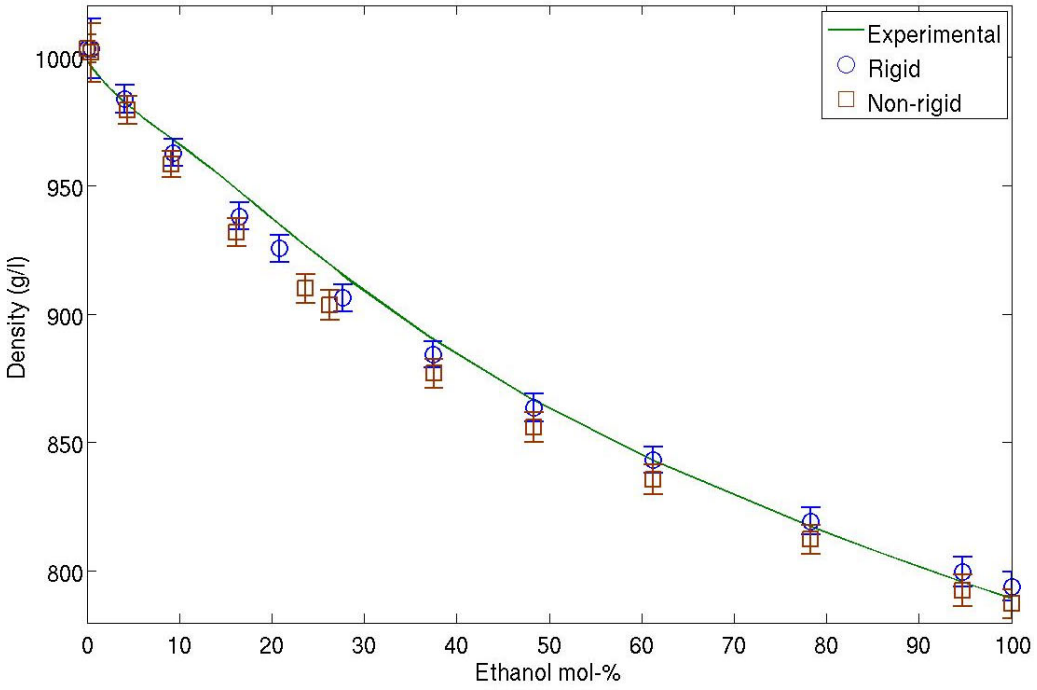

Figure 9: Density of the water-ethanol mixture with different ethanol concentrations. Rigid model (blue circles), non-rigid model (red squares) and experimentally obtained values (solid line) [60].

4 Results for molecular dynamics simulations

4.1 Density

In Fig. 9 are the densities obtained from the simulations. The experimental values [60] are also shown in the figure. In addition, the obtained densities are listed in Table 1. The values are average densities over the whole simulation and the errors are rms(root mean square)-fluctuations. Both the non-rigid and the rigid models give the trend in the density similar to the experimental value. Both models slightly underestimate the density, especially in concentrations between 10-40%. The non-rigid model underestimates the density to higher concentrations, up to 95%. The non-rigid model has a lower density than the rigid model in all concentrations, but overall the difference between the models is rather small. Wensink et. al. [12] calculated the density of water-ethanol mixtures using OPLS and TIP4P force fields. Also, in their study, the density was underestimated in lower concentrations, up to 40%.

4.2 Radial distribution functions

The studied radial distribution functions are distance distributions of water oxygen to water oxygen, OwOw, (Fig. 10), water oxygen to ethanol oxygen, OwOH, (Fig. 11) and ethanol oxygen to ethanol oxygen, OHOH, (Fig. 12). The RDF:s are averaged over the whole simulation run.

All the RDF:s show one sharp and high peak at around 2.8 Å and much smaller and smoother structures after that. After about 6 Å the radial distribution functions approach unity. The systematic effect is that the first peaks get higher and slightly narrower when ethanol concentration grows. It is notable that the shape of the radial distribution function is different for the three studied cases. To emphasize the difference in the first peak of the RDF, all three are plotted in the same figure (Fig. 13), which corresponds to a mixture of about 16% of ethanol. The basic features are the same for all the radial distribution functions, but the first peak is highest and narrowest for the water-water distribution and lowest and broadest for the ethanol-ethanol distribution, and the water-ethanol distribution lies in between. Altough the example is from the solution with 16% ethanol concentration, a similar trend applies for all concentrations.

As can be seen from Fig. 10-13 there is no significant difference between the non-rigid and rigid models. Some of the differences might be partly explained by the differences in the ethanol concentration which vary a little for the rigid and non-rigid models (in Fig. 10-13 or Table 1).

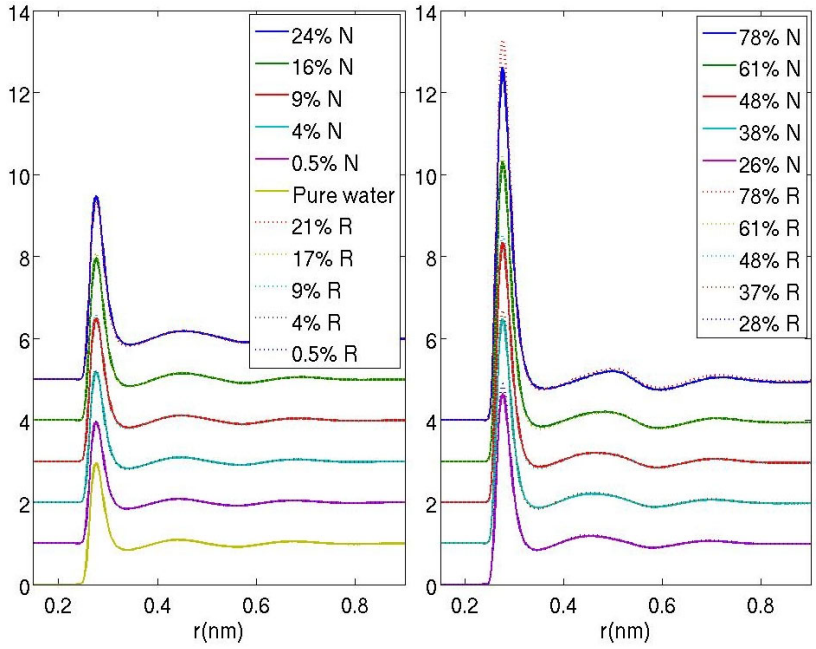

Figure 10: Radial distribution functions from water oxygen to water oxygen for different ethanol concentrations. Solid lines are for non-rigid ethanol (N) and dashed lines for rigid ethanol (R). Offsets are 1.0 units in the y-axis.

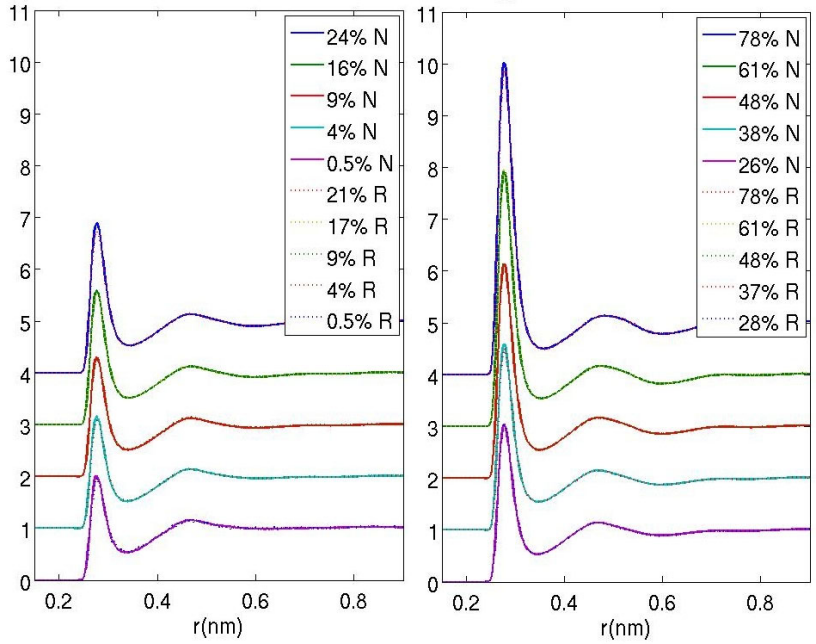

Figure 11: Radial distribution functions from water oxygen to ethanol oxygen for different ethanol concentrations. Solid lines are for non-rigid ethanol (N) and dashed lines for rigid ethanol (R). Offsets are 1.0 units in the y-axis.

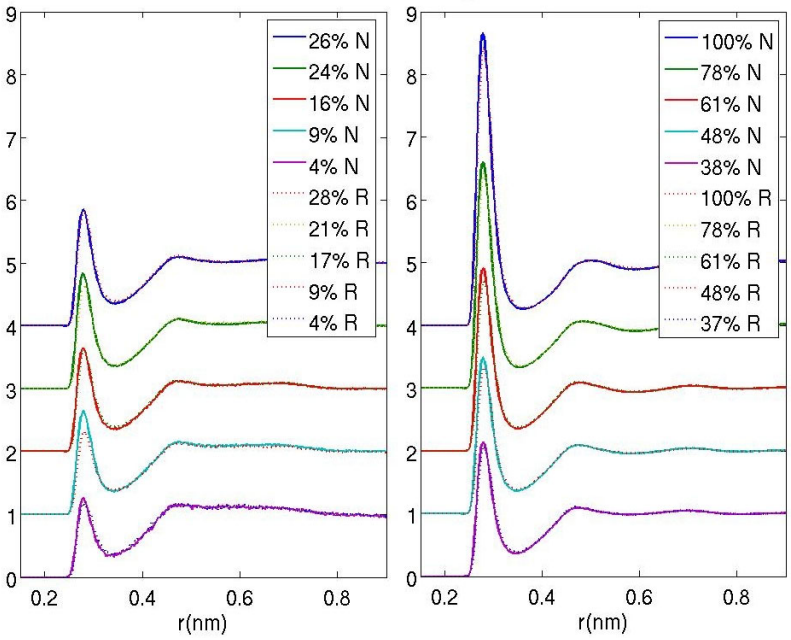

Figure 12: Radial distribution functions from ethanol oxygen to ethanol oxygen for different ethanol concentrations. Solid lines are for non-rigid ethanol (N) and dashed lines for rigid ethanol (R). Offsets are 1.0 units in the y-axis.

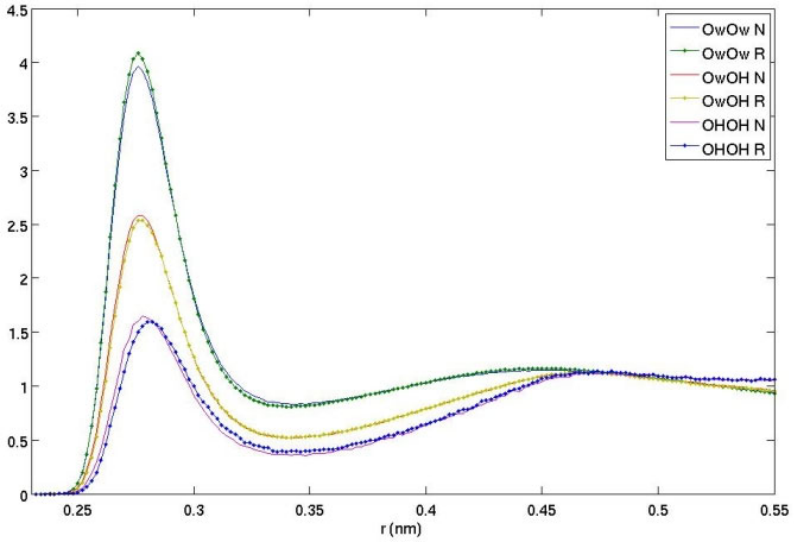

Figure 13: Radial distribution functions for a mixture with 16,2% of non-rigid ethanol (N) and 16,5% of rigid ethanol (R): water oxygen to water oxygen (OwOw), water oxygen to ethanol oxygen (OwOH) and ethanol oxygen to ethanol oxygen (OHOH).

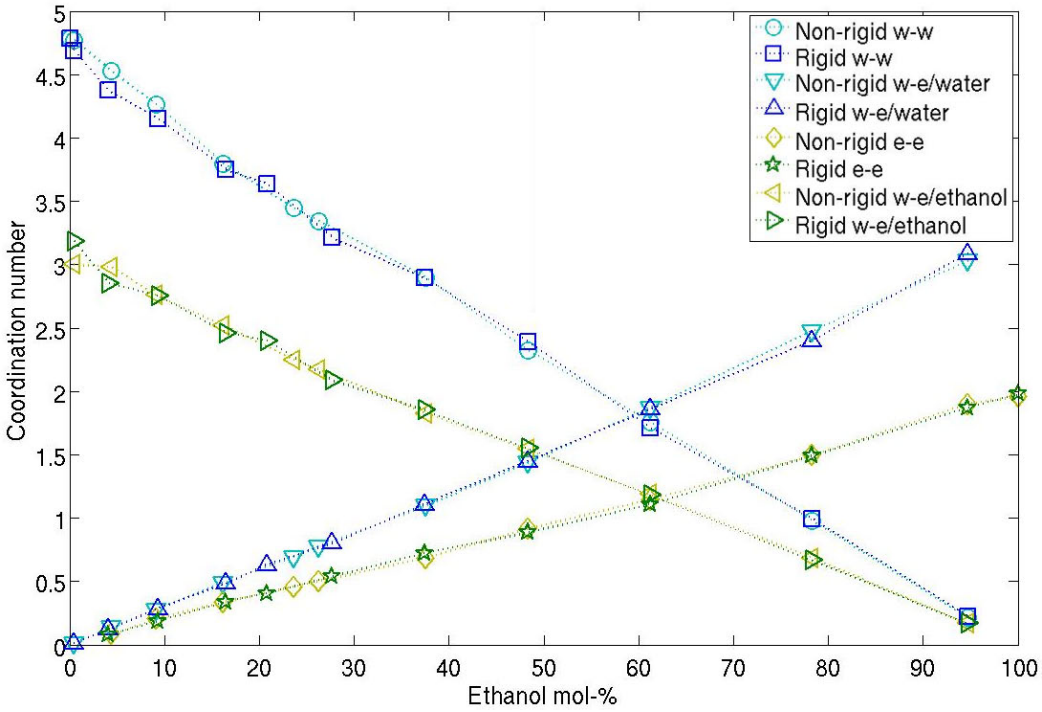

Figure 14: Coordination numbers for different ethanol concentrations. From top to bottom (small concentrations): water around water (w-w), water around ethanol (w-e/ethanol), ethanol around water (w-e/water) and ethanol around ethanol (e-e) coordination numbers. Note that the first point is for 0.5% of ethanol.

4.3 Coordination numbers

The studied coordination numbers are water around water (w-w), ethanol around ethanol (e-e), ethanol around water (w-e/water) and water around ethanol (w-e/ethanol). In Fig. 14 the coordination numbers for different ethanol concentrations are plotted.

It is observed that the change in the number of neighbours is almost linear. The ethanol molecules have overall less neighbours than the water molecules. Similarly as in the radial distribution functions there is only a small difference between the non-rigid and rigid models.

In this study the coordination number for pure water is 4.79 and for pure ethanol 1.96 and 1.97 for the non-rigid and rigid models, respectively. In a study of water-ethanol mixtures by Noskov et al. [13], where a polarizable force field for ethanol and SWM4-DP model for water was used, the results of coordination numbers were similar to this study. Noskov et al. found for pure water the coordination number of 4.63 and for pure ethanol 1.94, which both are close to the values obtained in this study,

which indicates that even with different models used in the calculation the amount of close neighbours for pure water and ethanol is constant.

4.4 Number of hydrogen bonds

In Fig. 15 the hydrogen bonds for a water molecule with different ethanol concentrations and in Fig. 16 the hydrogen bonds for an ethanol molecule with different ethanol concentrations are shown. In both figures the bonds with ethanol molecules and water molecules are plotted separately, as well as the total amount of bonds, which is the sum of the two. The linear lines represent the ideal mixing [13].

The number of hydrogen bonds for pure water is 3.50, and for non-rigid ethanol 1.82 and for rigid ethanol 1.80. These are close to the values received in other calculations. With the same hydrogen bond definition as used here, Zhang et al. [20] received 3.62 hydrogen bonds for pure water. With O-O distance less than 3.4 Å, van Erp et al. [19] received 3.4 hydrogen bonds for pure water. With a hydrogen bond definition of the H-O distance less than 2.4 Å, Zhang et al. received for pure water 3.04 hydrogen bonds and 1.69 hydrogen bonds for pure ethanol and Noskov et al. [13] 3.03 hydrogen bonds for pure water and 1.65 hydrogen bonds for pure ethanol. In a simulation calibrated against neutron diffraction data, Soper et al. [61] received 3.58 hydrogen bonds for pure water, with the definition of hydrogen bond of the O-O distance less than 3.5 Å.

The number of hydrogen bonds is smaller than the coordination number, which is expected, since not all the neighbouring molecules are hydrogen bonded to the central molecule. The difference between coordination number and number of hydrogen bonds is larger when there is more water in the mixture, than with more ethanol in the mixture. This means that the closest ethanol molecules are more likely to be hydrogen bonded with the centering molecule, whereas the closest water molecules are less likely to be bonded with the centering molecule.

Another interesting observation is that the total number of hydrogen bonds per water molecule and per ethanol molecule decreases as the ethanol fraction increases. For water the difference is about 15% and for ethanol 28%. The increase in ethanol hydrogen bonds when ethanol concentration decreases is therefore almost two times larger than the bonds water loses when ethanol concentration increases.

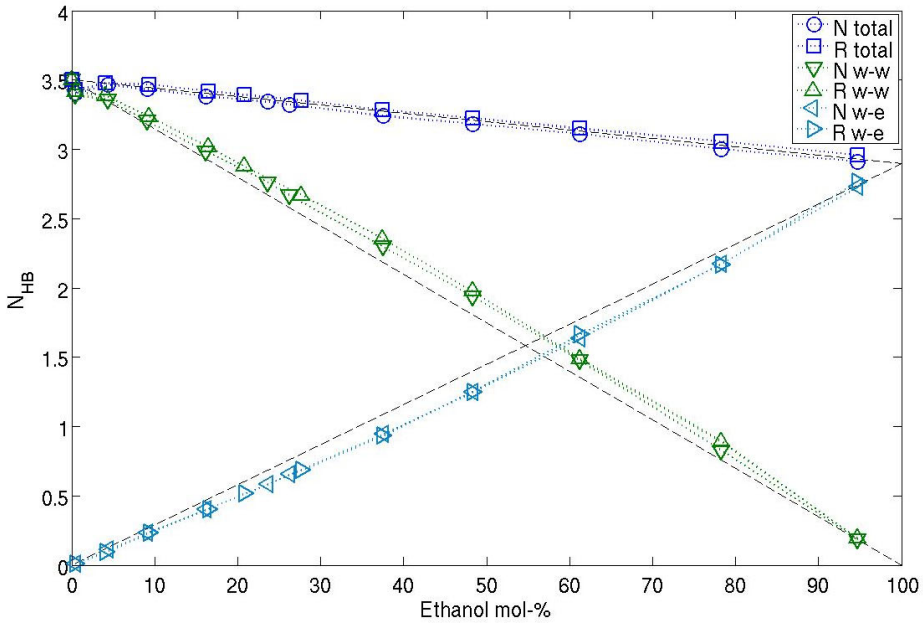

Figure 15: Number of hydrogen bonds for water for different ethanol concentrations. Blue for total amount of hydrogen bonds, green for bonds with water and light blue for bonds with ethanol. Black dashed lines are linear. Both rigid (R) and non-rigid (N) models.

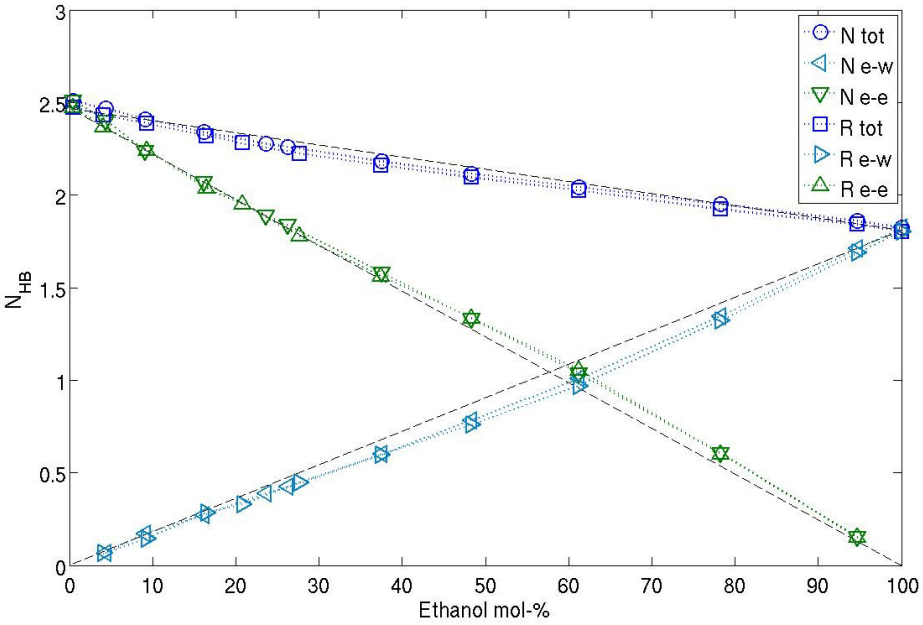

Figure 16: Number of hydrogen bonds for ethanol for different ethanol concentrations. Blue for total amount of hydrogen bonds, green for bonds with water and light blue for bonds with ethanol. Black dashed lines are linear. Both rigid (R) and non-rigid (N) models.

One notable feature in the hydrogen bonds is that the results are not linear as it would be for ideal mixing. There is positive excess for bonds with water (water-water, ethanol-water) and negative excess for bonds with ethanol (water-ethanol, ethanol-ethanol).

There is also a small difference between the ethanol models. The difference is most notable when both figures (Fig. 15 and 16) are compared. In Fig. 15 the total number of hydrogen bonds is slightly greater for rigid ethanol, and in Fig. 16 the total number of hydrogen bonds is slightly greater for non-rigid ethanol. This indicates that rigid ethanol bonds more with water and non-rigid ethanol more with ethanol.

4.5 Discussion

The good accordance with experimental values in densities of the water-ethanol mixtures gives a good starting point for further studies of the atomic-scale structure. The water model TIP4P and the ethanol model OPLS-AA are both fitted to give correct densities for pure liquids [40], and it is not surprising that they both together in mixtures give densities that are close to experimental values, although this is not always guaranteed. The small deviations from experimental values can be due to the use of cut-offs in the force fields. It is suggested that density is affected by the long-range interactions [12]. For pure water it was found that correlations exist at least up to 1.4 nm [62], which suggests that the cut-offs used in this study are insufficient to produce correct long-range correlations.

The smaller density of non-rigid ethanol model in all concentrations can be understood by the intramolecular vibrations, which make the molecules on the average a bit larger and so the density is smaller. The rigid model, which vibrates less, can be packed more densely. However, overall the difference between the models is small.

The radial distribution functions show that there is a clear first and second hydration shell in the mixtures. Especially the first hydration shell (the first peak) is very clear. The sharpening of the peak when ethanol concentration grows indicates that the mixture is more structured when more ethanol is involved. Also, when comparing different oxygen to oxygen radial distribution functions, the peak is higher and narrower for the water to water distribution than for the water to ethanol or ethanol to ethanol distributions, which indicates that water has more ordered structure in its nearest neighbouring area.

Although the coordination number changes linearly as the ethanol concentration grows, the behaviour of the number of hydrogen bonds is not linear. The positive excess bonding with water and negative excess bonding with ethanol indicates that water is a better solvent than ethanol, since bonds are preferably formed with water. The same conclusion can be made from that ethanol gains totally almost 30% of more bonds when water concentration increases, but water only loses totally about 15% of bonds when water concentration decreases. The differences between the ethanol models in the total number of hydrogen bonds indicate that when rigid ethanol is mixed with water, water bonds more with water, and with non-rigid ethanol water bonds less with water. However the difference in bonding is overall very small.

Some differences between the non-rigid and rigid ethanol models can be found, but they are rather small. This indicates that the flexibility in bond lengths affects the simulation, but these studied quantities gave only a small hint of the difference and more accurate methods are needed to fully investigate the deviation between the ethanol models.

However, since the differences between the models are so small, it can be said that in the MD simulations the non-rigid and rigid ethanol model give the same results in these studied quantities. If thought time-wise, it takes twice the time to run the simulations for non-rigid ethanol, since the time step is smaller than for the rigid model. That way, if fast simulations are needed, it is always reasonable to use rigid ethanol, since the time step can be longer.

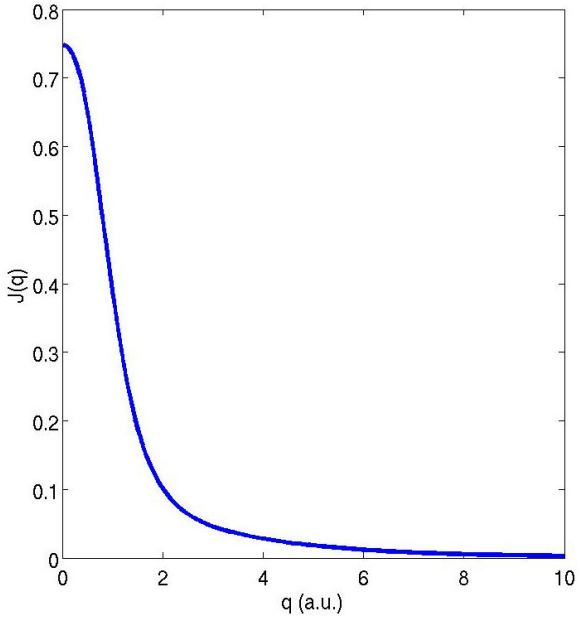

Figure 17: Sample Compton profile for pure water. The profile is an average obtained from 80 snapshots.

5 Results for Compton profiles

5.1 Compton profiles for pure liquids

To start the study of the Compton profiles, it is useful to first discuss the estimation of the errors in the results. The first step is to study pure liquids. The pure liquid Compton profiles are then later on used in the calculation of difference Compton profiles.

Fig. 17 shows an example of a computationally obtained Compton profile. It is an average of 80 Compton profiles for pure water. All the Compton profiles (including the mixture profiles) have a similar shape.

The errors in the Compton profiles of pure liquids are estimated by the bootstrap method (Section 3.3). For pure water there were 80 clusters, from which 1000 bootstrap profiles and the average profile were calculated. In Fig. 18 are plotted the difference between the bootstrap profiles and the average profile. This kind of plot is similar also for pure ethanol and to the mixtures.

From the bootstrap profiles a standard deviation error profile can be calculated. For the error estimate, this standard deviation in the bootstrap profile is used. As a

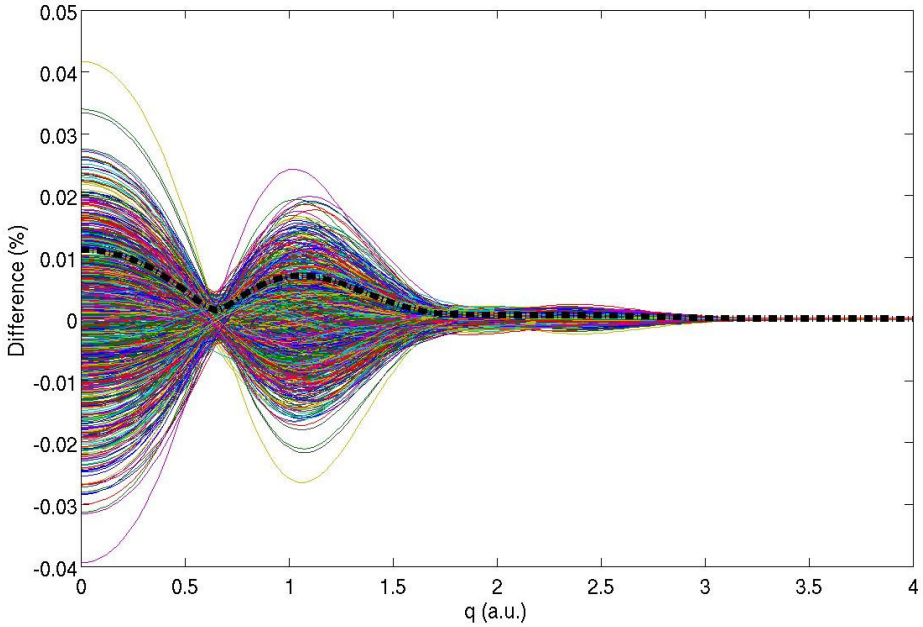

Figure 18: Example of 1000 bootstrap profiles compared to average profile from 80 clusters of pure water. The black dashed line is the standard error deviation profile.

Table 2: Statistical errors for the Compton profiles of pure liquids calculated with the bootstrap method. The substance/model, the amount of clusters, the number of bootstraps, and the error at $J(0)$ is given.

	Amount of clusters	**Amount of bootstraps**	**Error (%-units) at $J(0)$**
Water	80	1000	0.0108
Rigid ethanol	100	1000	0.0059
Non-rigid ethanol	100	1000	0.0141

conservative estimate, the biggest difference to zero (typically occurring at $q = 0$ a.u.) in standard deviation profile is then taken to be the error in the Compton profile. For pure water the error in the Compton profile is 0.0108%-units. This is also listed in Table 2.

The errors for pure ethanol are calculated similarly for both the non-rigid and rigid models. There were 100 clusters for both non-rigid and rigid models, from which 1000 bootstrap profiles and the average profile were calculated. For rigid ethanol the error is 0.0059%-units and for non-rigid the error is 0.0141%-units. These results are also listed in Table 2.

A notable thing in the Compton profiles for pure liquids is that the error estimate is over two times larger for the non-rigid than for the rigid ethanol. The error in

the water Compton profile is between the different ethanol models, but there are also less clusters for water, which affects the error estimate calculated with the bootstrap method.

5.2 Difference Compton profiles

In this section the difference Compton profiles for water-ethanol mixtures are studied. Subsequently the errors are estimated.

The difference Compton profiles were studied for mixtures with 4.4%, 17% and 95% rigid ethanol concentration (in Fig. 19) and for mixtures with 4.4%, 18% and 93% of non-rigid ethanol concentration (in Fig. 20). For all the calculations 100 clusters have been used.

For rigid ethanol all the difference Compton profiles look rather similar. They all begin from the negative side and then go to the positive side before going to zero. Although there are differences between the curves, no consistent trend can be observed. For the 17% concentration the deviations from zero are largest, and for the 95% concentration smallest.

The case is very different for non-rigid ethanol. In this case the difference Compton profiles are dissimilar. For the 4.4% concentration the difference Compton profile begins from the negative side and then goes to positive side and back to negative side before going to zero with some fluctuation. At the 18% concentration the difference Compton profile has similar shape as the 4.4% concentration, but it begins from the positive side, and there is still a peak at the positive side. Both of the lower concentration cases are overall very close to zero (the deviation from zero is less then 0.01%-units). The 93% concentration is completely different shape-wise. It begins from the positive side and then goes to the negative side before going to zero. There are no peaks at the positive side.

In order to emphasize the differences between the ethanol models, Fig. 21 shows the difference Compton profiles for both models with the same concentration in the same figure. The non-rigid and rigid models give rather similar difference Compton profiles when the concentration is small, but with larger concentrations the difference Compton profiles are quite distinct. For the 17-18% concentration the non-rigid model

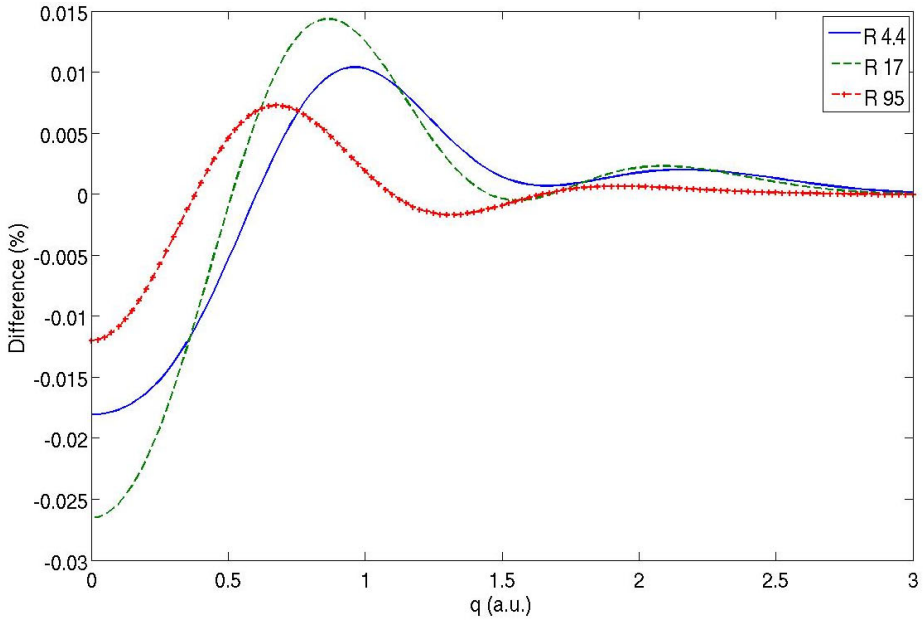

Figure 19: Difference Compton profiles of mixtures with 4.4%, 17% and 95% of rigid (R) ethanol.

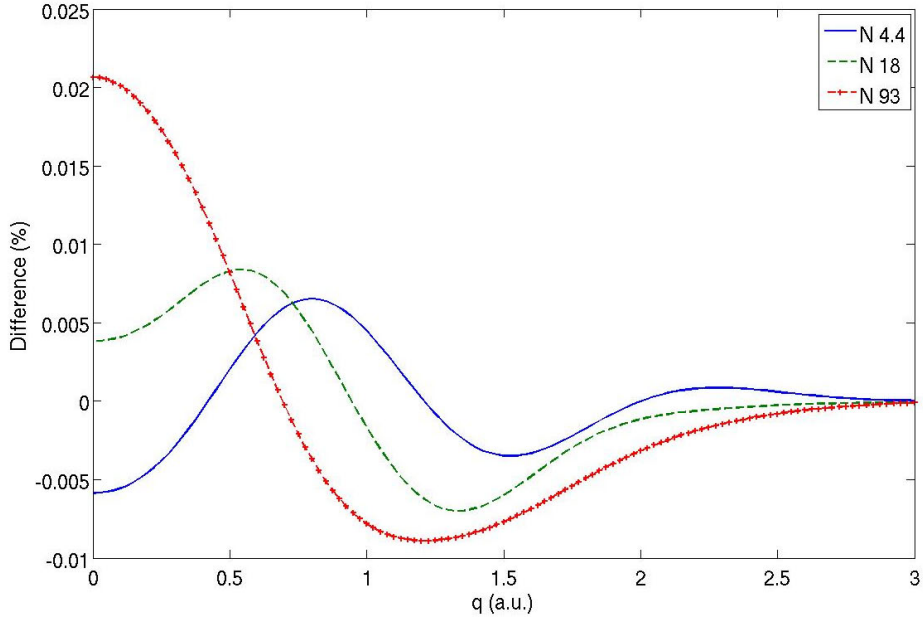

Figure 20: Difference Compton profiles of mixtures with 4.4%, 18% and 93% of non-rigid (N) ethanol.

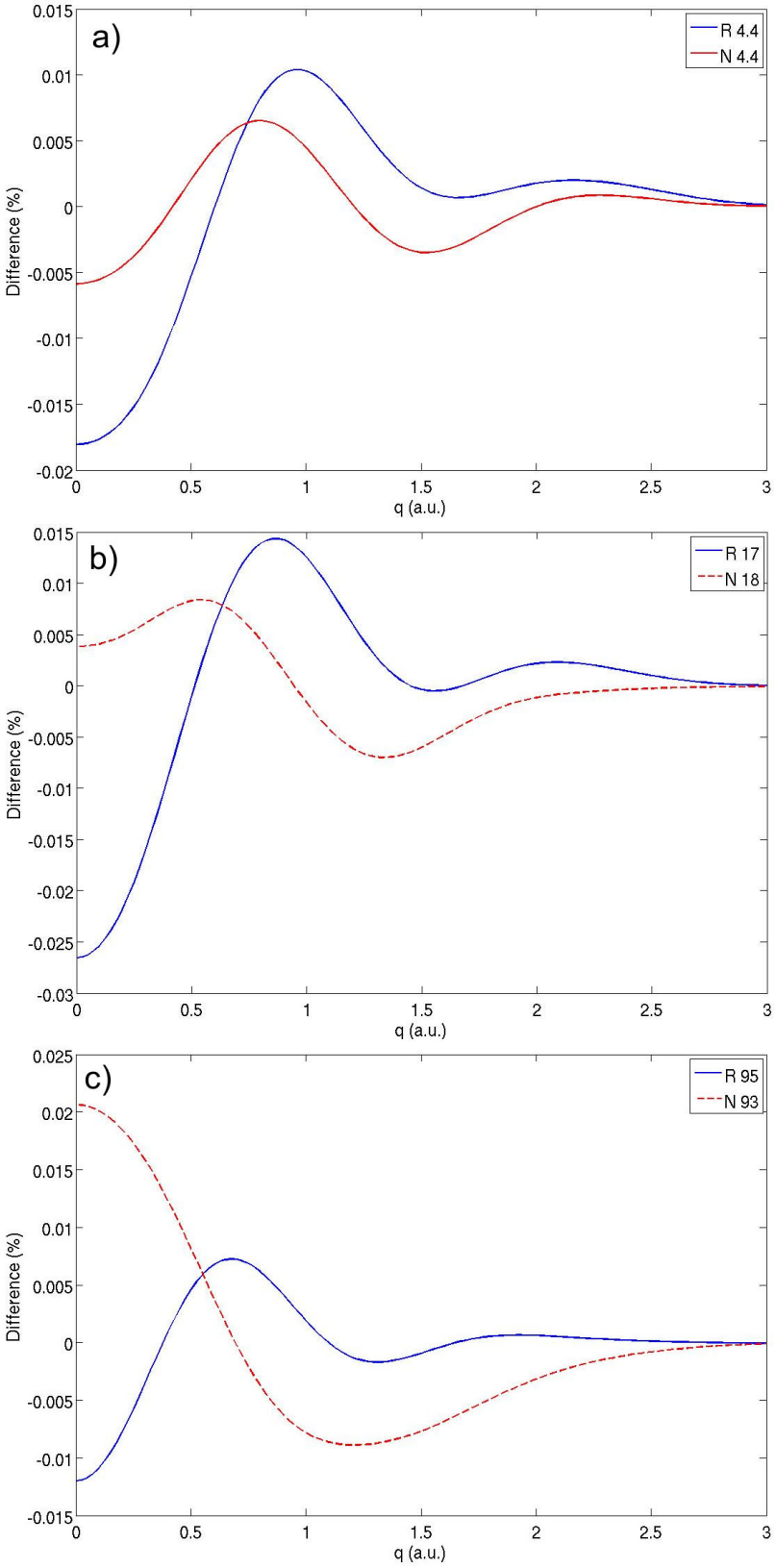

Figure 21: Difference Compton profiles for mixtures with a) 4.4% rigid and non-rigid ethanol, b) 17% of rigid and 18% of non-rigid ethanol and c) 95% of rigid and 93% of non-rigid ethanol. Solid lines for rigid (R) model and dashed lines for non-rigid (N) models.

Table 3: Error estimates for the difference Compton profiles calculated with the bootstrap method. Model and concentration of ethanol, Compton profile error, average ethanol concentration in the bootstrap profiles and the error in the concentration of the bootstrap profiles are given.

	Error (%-units) at $J(0)$	**Bootstrap average concentration**	**Bootstrap concentration error**
Rigid 4.4%	0.0094	4.09	0.29
Rigid 17%	0.0095	17.32	0.67
Rigid 95%	0.0071	95.21	0.489
Non-rigid 4.4%	0.0101	4.72	0.34
Non-rigid 18%	0.0109	18.03	0.74
Non-rigid 93%	0.0145	93.07	0.49

gives a profile very close to zero, but the rigid model has the largest deviations from zero. Also the rigid model begins from the negative side, but the non-rigid from the positive side. Otherwise the shapes are similar with the peaks in the positive side. The biggest difference is in the 93-95% concentration. There the non-rigid profile is of completely different shape and the profiles begin from the different sides of zero, the rigid model from the negative and the non-rigid from the positive.

The error estimation is a bit more challenging for difference Compton profiles, since in the bootstrap method the concentration for the different bootstrap profiles varies. The error estimates are made for all the concentrations and for the both ethanol models. In all calculations there has been 100 clusters and 1000 bootstraps. The bootstrap profile differences to the average profile difference look similar to the pure water case (Fig. 18).

The Compton profile error estimation is made the same way as for the pure liquids using the bootstrap method. In Table 3 are listed the Compton profile errors for the different ethanol concentrations and models. In this error estimate it is assumed that there is no error in the pure liquid profiles. Notably, the concentrations obtained by the bootstrap method are very close to the actual concentration. There is still some additional error occurring from the fact that the profiles are weighted according to the concentration. For the rigid model, the Compton profile error decreases as the concentration increases, but for the non-rigid model the situation is completely opposite. The difference between the models increases when there is more ethanol in the mixture.

5.3 Discussion

When studying the difference Compton profiles for binary liquids, there are three conceivable scenarios. Firstly, if the liquids do not mix or the mixing structure is similar to the bulk structure of pure liquids, the difference Compton profiles should be zero. Secondly, if the mixing is similar in all the concentrations, the difference Compton profiles should be the same. Finally, if the mixing is not similar in different concentrations, the difference Compton profiles should be distinguishable.

It is obvious that in this study two different scenarios are seen depending on the ethanol model. For rigid model the difference Compton profiles are very similar. The shape of the profiles is uniform, although the magnitude varies a little. This would mean that for the rigid ethanol the water-ethanol mixing is similar in all concentrations. The same does not apply to the non-rigid ethanol. There is a difference between the low and the high ethanol concentration in the difference Compton profiles. Both the shape and the magnitude of the difference Compton profile changes when there is more ethanol in the mixture. This would indicate that for the non-rigid ethanol the mixing changes when the ethanol concentration grows. For high ethanol concentration the structure created in the mixing process is different from the lower ethanol concentrations. Also for the non-rigid ethanol the low-concentration difference Compton profiles are very close to zero, which would indicate that the hydrogen bonded network of molecules is similar to the bulk liquids i.e. adding small amount of (non-rigid) ethanol to the mixture does not affect the structure of water.

The systematic error in the Compton profiles comes from several sources. Firstly, the used molecular configurations are made without using quantum mechanics. This can lead to the molecular structures being incorrect. Secondly, there are several approximations in the calculation of the Compton profiles. Also, the used cluster size, which is far from bulk material surely affects the result.

The statistical error occurs from the limited number of chosen clusters. From the Compton profiles of pure liquids the errors for the profiles are estimated to be of the order of 0.01%-units. However, these are ignored in the error estimation of the difference Compton profiles, for which the errors are estimated to be of the same order. The smallest difference Compton profile deviates from zero about 0.01%-units and the biggest deviation is only 0.025%-units.

Due to these uncertainties, which are of the order of typical experimental inaccuracy, the results of this study can be considered preliminary. To confirm these results more

statistics is needed. Thus for further and more accurate studies more clusters must be calculated.

The difference in statistical fluctuation in the Compton profiles when using the rigid and non-rigid models for ethanol is remarkable. The error is almost 1,5 times larger for the non-rigid ethanol. Compton scattering is highly dependent on the bond length [63]. Since the non-rigid ethanol, without restrictions in the bond length, has a greater amount of fluctuations in the intramolecular bond lengths, the atomic structures vary more for the non-rigid ethanol than for the rigid ethanol. This causes more changes in the Compton profiles and thus gives larger statistical fluctuation.

Also the difference in the Compton profiles of the mixtures with higher ethanol concentration for the non-rigid ethanol can be explained by the fluctuations in the bond lengths. This does not show at low ethanol concentrations, where the non-rigid and rigid cases are similar, because there water is the dominant substance in the mixture, but when the ethanol concentration grows, and ethanol becomes dominant, the bond length fluctuations start to affect the Compton profile significantly.

However, there are not many data points for the mixture Compton profiles as only three concentrations are studied. To be certain of the higher ethanol concentration behaviour, more than one high ethanol concentration mixture should be studied. Also the intermediate concentrations are worth studying, because that would bring more information on the mixing behaviour.

6 Conclusions

Combining simulation results from classical molecular dynamics with x-ray Compton scattering analysis new information will be received on structural properties of liquids. Although the MD method used in this study does not include quantum mechanical phenomena explicitly, by using Compton scattering as a further analysis tool, new insight into the structures created by the molecular dynamics method can be found. From this study some important findings can be extracted for both the structure of water-ethanol mixtures and for the molecular dynamics method itself.

Firstly, from the MD simulations, the mixture densities were studied. The results slightly underestimated the experimental values. Then the microscopic, atomic-scale properties of the mixtures were studied. The radial distribution functions suggest that there would be more ordered structure when ethanol concentration grows, and water has more ordered structure in mixtures than ethanol. The study of hydrogen bonding shows that there is excess bonding with water and deficient bonding with ethanol in the mixture. Both ethanol and water form bonds preferably with water, which shows that water is a better solvent than ethanol, which is expected, since water is known for its solvation properties. These quantities showed only little or no difference between the non-rigid and the rigid ethanol model.

When Compton scattering was used to focus on the atomic-scale and electronic properties, there was a notable difference between the models. This indicates that the choice of the force field model is critical in microscopic structure formation in molecular dynamics. Even though some of the MD results could be considered similar for both models, others hinted that there is some differences between the models. This was then clearly visible with the Compton calculations, since the Compton profiles depend not only on the local intermolecular geometries but also on the intramolecular bond lengths [63]. Firstly, there was a clear difference between the ethanol models in difference Compton profiles for high ethanol concentration. Secondly, the size of the errors in pure ethanol and at high ethanol concentration showed differences between the models.

It also seems that the rigid ethanol mixes similarly to water in all concentrations, but non-rigid does not. The models mix similarly in low ethanol concentrations, where water is the dominant substance, but when there is more ethanol in the mixture, non-rigid ethanol mixes differently than rigid and the low concentration cases. This

would mean that something is different in the mixing process for non-rigid ethanol with high ethanol concentrations.

However these results are preliminary, since the statistics are not sufficient enough, which is shown in the error scales for the Compton profiles. This means that more than 100 clusters are needed for binary liquid mixtures to get good enough statistics.

For non-rigid ethanol the Compton profile errors were notably larger than for rigid ethanol. This means that for non-rigid ethanol even more clusters are needed to get sufficient statistics, and the time needed for calculations grows large. With one MD simulation run taking several days on desktop computer, and every Compton profile calculation for one cluster taking on average three hours, the use of non-rigid model will be time-consuming. With the rigid model the MD simulation is faster and less clusters are needed, so the whole process is much faster. So if only low concentration calculations are needed, using rigid ethanol model is more sensible.

After getting more statistics to the present results, they can be compared to experimental data. Then valuable information about the models and the mixing process is obtainable, which will show whether the MD simulations can produce correct microscopic structures. If some of the MD structure data will fit the experimental data, it is also possible to see which model produces the correct structure. This will be a very interesting task in the future.

With adding more concentrations in the Compton profile calculations, more information about the mixing behaviour can be obtained. It will be possible to see at which concentration the mixture Compton profiles of non-rigid ethanol changes and whether the change is gradual or sudden. It will be interesting to see if the rigid model mixes in fact similarly for all the concentrations, including the intermediate ones. This is another important subject to be studied in the future.

All in all, with furthering this study a bit and doing the experimental work, it is possible to obtain valuable information about binary liquids. It is possible to see whether classical molecular dynamics is enough to produce correct microscopic structures or if there is something else needed for accuracy. If the method gives proper results, it can be then applied to other systems. Also the amount of rigidity needed in the force fields for the MD simulation can be solved.

References

[1] S. Dixit, J. Crain, W. C. K. Poon, J. L. Finney and A. K. Soper, Nature, **416** 829 (2002).

[2] C. H. Patrick, *Alcohol, culture and society* (Duke University Press, Durham, 1952).

[3] R. H. Blum and Assosiates, *Society and Drugs* (Jossey Bass, San Francisco, 1969).

[4] S. P. Lucia, *A history of wine as Therapy* (J. B. Lippincott, Philadelphia, 1963).

[5] B. L. Vallee, Scientific American, **278** 85 (1998).

[6] R. L. Myers, *The 100 most important chemical compounds: a reference guide* (Greenwood Publishing Group, Santa Barbara, 2007).

[7] T. Babor, *Alcohol: Customs and Rituals* (Chelsea House, New York, 1986).

[8] G. McDonnel and A. D. Russell, Clin. Microbiol. Rev., **12** 147 (1999).

[9] M. Windholz, *The Merck index: an encyclopedia of chemicals and drugs* (Rahway, New York, 1976).

[10] L. R. Lynd, Annu. Rev. Energy Environ., **21** 403 (1996).

[11] A. Nose, T. Hamasaki, M. Hojo, R. Kato, K. Uehara and T. Ueda, J. Agric. Food Chem., **53** 7074 (2005).

[12] E. J. W. Wensink, A. C. Hoffmann, P. J. van Maaren and D. van der Spoel, J. Chem. Phys., **119** 7308 (2003).

[13] S. Y. Noskov, G. Lamoureux and B. Roux, J. Phys. Chem. B, **109** 6705 (2005).

[14] F. Müeller-Plathe, Mol. Simul., **18** 133 (1996).

[15] V. N. Levchuk, I. I. Sheykhet and B. Y. Simkin, Chem. Phys. Lett., **185** 339 (1991).

[16] J. Fidler and P. M. Roger, J. Phys. Chem. B, **103** 7695 (1999).

[17] M. Tarek, D. J. Tobias and M. L. Klein, Physica A, **231** 117 (1996).

[18] L. Zhang, Q. Wang, Y.-L. Liu and L.-Z. Zhang, J. Chem. Phys., **125** 104502 (2006).

[19] T. S. van Erp and E. J. Meijer, J. Chem. Phys., **118** 8831 (2003).

[20] Y. Zhang and S. Patel, J. Phys. Chem. B, **113** 767 (2009).

[21] J. von Boehm, *Molekyylidynamiikkamenetelmä* (Otatieto, Helsinki, 2000).

[22] B. Smit and D. Frenkel, *Understanding Molecular Simulation, From Algorithms to Applications* (Academic Press, Orlando, 2002).

[23] B. J. Alder and T. E. Wainwright, J. Chem. Phys., **27** 1208 (1957).

[24] B. J. Alder and T. E. Wainwright, J. Chem. Phys., **31** 459 (1959).

[25] F. H. Stillinger and A. Rahman, J. Chem. Phys., **60** 1545 (1974).

[26] M. J. Cooper, P. E. Mijnarends, N. Shiotani, N. Sakai and A. Bansil, *X-ray Compton Scattering* (Oxford University Press Inc., New York, 2004).

[27] M. Hakala, K. Nygård, S. Manninen, L. G. M. Pettersson and K. Hämäläinen, Phys. Rev. B, **73** 035432 (2006).

[28] M. Hakala, K. Nygård, J. Vaara, M. Itou, Y. Sakurai and K. Hämäläinen, J. Chem. Phys., **130** 125413 (2009).

[29] Editor B. Williams, *Compton Scattering* (McGraw-Hill Inc., Great Britain, 1977).

[30] J. K. Laukkanen, *Advanced experimental methods in Compton scattering spectroscopy*, Ph.D. thesis, University of Helsinki (2000).

[31] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen and H. J. C. Berendsen, *Gromacs User Manual version 4.0* (www.gromacs.org, 2005).

[32] RCSB protein data bank, http://www.rcsb.org/pdb/home/home.do (2009).

[33] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1992).

[34] R. Kubo, *Statistical mechanics* (North-Holland publishing company, Amsterdam, 1965).

[35] D. C. Rapaport, Computer Physics Communications, **174** 521 (2006).

[36] L. Verlet, Phys. Rev., **159** 98 (1967).

[37] A. R. Leach, *Molecular modelling: principles and applications* (Pearson Education Limited, Dorchester, 2001).

[38] W. L. Jorgensen, J. Phys. Chem., **87** 5304 (1983).

[39] J. Delhommelle and P. Millie, Molecular Physics, **99** 619 (2001).

[40] I. Juurinen, *Molekyylimekaniikan vuorovaikutusmallit*, University of Helsinki (2009). Bachelor thesis.

[41] S. Brandt, *Statistical and Computational Methods in Data Analysis* (North-Holland publishing company, Amsterdam, 1970).

[42] M. C. Payne, M. P. Teter, D. C. Allan, T.A. Arias and J. D. Joannopuolos, Rev. Mod. Phys., **64** 1045 (1992).

[43] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola and J. R. Haak, J. Chem. Phys., **81** 3684 (1984).

[44] A. Kuronen, *Introduction to atomistic simulations, lecture notes*, University of Helsinki (2008).

[45] F. Mandl, *Statistical physics* (John Wiley & sons Ltd., Bath, 1980).

[46] L. H. Schwartz and J. B. Cohen, *Diffraction from materials* (Academic Press Inc., New York, 1977).

[47] A. Rahman and F. H. Stillinger, J. Chem. Phys., **55** 3336 (1971).

[48] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, J. Chem. Phys., **79** 926 (1983).

[49] W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, J. Am. Chem. Soc., **118** 11225 (1996).

[50] B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, J. Comput. Chem., **18** 1463 (1997).

[51] H. J. C. Berendsen, D. van der Spoel and R. van Drunen, Comp. Phys. Comm., **91** 43 (1995).

[52] R. J. Weiss, *X-ray Determination of electron Distributions* (North-Holland publishing company, Amsterdam, 1966).

[53] K. Nygård, *Local structure of water studied by Compton scattering*, Ph.D. thesis, University of Helsinki (2007).

[54] J. L. Heilbron, *Historical studies in the theory of atomic structure* (Arno press, A New York Times Company, New York, 1981).

[55] L. Van Hove, Phys. Rev., **95** 249 (1954).

[56] M. A. L. Marques, C. A. Ullrich, F. Nogueira, A. Rubio, K. Burke and E. K. U. Gross, *Time-Dependent Density Functional Theory* (Springer, Berlin Heidelberg, 2006).

[57] M. J. Cooper, Rep. Prog. Phys., **48** 415 (1985).

[58] K. Hermann and L. G. M. Pettersson, *StoBe-deMon Version 3.0*, http://www.fhi-berlin.mpg.de/ hermann/StoBe/index.html (2009).

[59] B. Efron and R. J. Tibshirami, *An introduction to the Bootstrap* (Chapmann & Hall Inc., London, 1993).

[60] Editor D. R. Lide, *CRC, Handbook of Chemistry and Physics, 82nd edition* (CRC Press LLC, Boca Raton, Florida, 2001).

[61] A. K. Soper, F. Bruni and M. A. Ricci, J. Chem. Phys., **106** 247 (1997).

[62] C. Chipot, C. Milot, B. Maigret and P. A. Kollman, J. Chem. Phys., **103** 7953 (1994).

[63] M. Hakala, S. Huotari, K. Hämäläinen, S. Manninen, Ph. Wernet, A. Nilsson and L. G. M. Pettersson, Phys. Rev. B, **70** 125413 (2004).