TAMPERE UNIVERSITY OF TECHNOLOGY

EERO HYTÖNEN

LIQUID DYNAMICS IN COMPUTER SIMULATIONS

Master of Science Thesis

Examiner: Ilpo Vattulainen
Examiner and topic approved in
the Science and Environmental
Faculty council meeting 04.06.2011

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Teknis-luonnontieteellinen koulutusohjelma
EERO HYTÖNEN: Nesteiden dynamiikkaa tietokonesimulaatioissa
Diplomityö, 64 sivua
Kesäkuu 2013
Pääaine: Teknillinen fysiikka
Tarkastaja: Professori Ilpo Vattulainen
Avainsanat: Biologinen fysiikka, diffuusio, viskositeetti, molekyylidynamiikka, karkeistetut mallit, vesi

Tietokoneiden laskentatehon eksponentiaalinen kasvu on vahvistanut laskennallisen tieteen merkitystä ja mahdollistanut yhä monimutkaisempiin ja laajempiin ongelmiin tarttumisen. Fysiikan ja biologian rajapinnassa toimiville tieteenaloille laskennallinen tutkimus on arvokas työkalu tutkittavien ilmiöiden luonteen vuoksi. Laskentatehon kasvun myötä mallinnettavien systeemien kompleksisuus on kasvanut samaan tahtiin, mikä on luonut tarpeen joustavien ja riittävän yksinkertaisten molekyylimallien kehittämiselle.

Biologisessa fysiikassa eräs käytetyimmistä laskennallisista tutkimusmenetelmistä on molekyylidynamiikka, jossa ratkotaan Newtonin liikeyhtälöitä ajan suhteen ja tuloksena saadaan atomien sijainnit ajan funktiona. Molekyylidynamiikkasimulaatiot vaativat erittäin paljon laskentaa, mikä rajoittaa niiden käytettävyyttä. Eräs ratkaisu laskennan keventämiseksi on molekyylimallien yksityiskohtien vähentäminen eli karkeistaminen. Karkeistamisessa useita atomeja yhdistetään yhdeksi partikkeliksi, jonka ominaisuudet vastaavat keskimäärin yhdistettyjen atomien ominaisuuksia. Kun simulaation muuttujien lukumäärä pienenee, laskenta kevenee ja simulaatio nopeutuu. Toinen simulaatiota nopeuttava tekijä on karkeistettujen molekyylien väliset yksinkertaistetut vuorovaikutukset, jolloin mikrotilojen tutkiminen on nopeampaa. Näistä johtuen karkeistetuilla malleilla tehtyjen si-mulaatioiden aikaevoluutio on nopeampaa kuin atomistisilla malleilla. Tutkittaessa aikaevoluution nopeutumista, havaitaan vain molempien nopeuttavien tekijöiden yhteisvaikutus. Jotta aikaskaalan eroavaisuuksia pystyy tutkimaan, tulee molempien nopeutumiskertoimien vaikutus tuntea erikseen. Karkeistettujen mallien rakenteellisia ja termodynaamisia ominaisuuksia on tutkittu paljon ja niiden on todettu jäljittelevät hyvin atomitason malleja, mutta dynaamisia eli ajasta riippuvia ominaisuuksia ei ole tutkittu yhtä tarkasti.

Tässä työssä tavoitteena on selvittää kuinka hyvin karkeistetut vesimallit täsmäävät oikean veden ja atomistisen veden dynaamisia ominaisuuksia, sekä tutkia eri dynaamisten ominaisuuksien aikaskaalojen eroja. Karkeistettujen ja atomististen nesteiden dynaamisten ominaisuuksien vertailussa käytetään vettä mallisysteeminä, ja

tutkitaan itseisdiffuusiota, kollektiivista diffuusiota sekä viskositeettia lämpötiloissa 300–350 K. Karkeistettuina malleina on käytössä niin sanottu Martini-malli ja polarisoituva Martini-malli, joita verrataan atomitason SPC-vesimalliin. Karkeistetuksi malliksi valittiin Martini ja atomistiseksi malliksi SPC, koska molemmat ovat laajasti käytettyjä ja hyväksi todettuja malleja. Simulaatiot tehtiin GROMACS 4.5.1-simulaatio-ohjelmistolla ja viskositeetin sekä kollektiivisen diffusion analysoinnit itsetehdyllä koodilla.

Tulokset osoittavat, etteivät aikaskaalat ole samoja eri dynaamisille ominaisuuksille, mutta ne ovat kuitenkin samaa suuruusluokkaa. Martini-mallin itseisdiffuusion tiedetään olevan neljä kertaa todellista diffuusiota nopeampaa, ja tämä on otettu yleiseksi nyrkkisäännöksi dynamiikan nopeutumiselle eri ilmiöissä. Havaitut nopeutumiskertoimet Martini-mallille ovat 2 (leikkausviskositeetti), 4 (itseisdiffuusio) ja 4 (kollektiivinen diffusio). Polarisoituvalle Martini-mallille luvut ovat 1,2, 5 ja 6. Kollektiiviselle diffuusiotekijälle ei laskentamenetelmästä johtuen saada yksikäsitteistä arvoa.

Karkeistettujen simulaatioiden aikaskaalan tulkinta ei siis ole suoraviivaista, sillä samalle mallille saadaan erilaisia nopeutumiskertoimia riippuen käytetystä dynaamisesta ominaisuudesta. Erot ovat kuitenkin maltillisia ja yleisesti ottaen tulokset tukevat Martini mallin semi-kvalitatiivista luonnetta. Aikaskaalan nopeutumisessa voidaan jatkossakin käyttää ensimmäsenä approksimaationa kerrointa neljä.

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Science and Technology
EERO HYTÖNEN : Liquid Dynamics in Computer Simulations
Master of Science Thesis, 64 pages
June 2013
Major: Engineering Physics
Examiner: Professor Ilpo Vattulainen
Keywords: Biological physics, diffusion, viscosity, molecular dynamics, coarse-graining

Coarse-grained models have become commonplace in computational biophysics research for their obvious benefits. They enable simulations of larger systems and longer time scales compared to atom-level models with reasonable accuracy and computational load. Coarse-grained models are built from interaction sites which represent a group of atoms and aim to describe the average properties of the underlying atomistic groups. The decreased computational load is achieved by having simplified interactions and fewer interacting particles. The simplified interactions and the coarse-graining process both create a speed up compared to atomistic models which results in faster time evolution in the simulation.

In this thesis the aim is to study how the dynamic properties of coarse grained descriptions of water correspond to real water and atomistic fine grained water, and also to see if there are differences in the time scales between different dynamic properties. The dynamic properties of liquids are compared through Molecular Dynamics simulations using water as the model system in the temperature range 300–350 K. Molecular Dynamics technique solves the Newton's equations of motion over discrete time steps. The SPC force field is chosen as the fine-grained model and the coarse-grained models are based on Martini and polarizable Martini force fields. Viscosity and self-diffusion are calculated using standard methods whereas collective diffusion is determined from the decay of the density fluctuation autocorrelation function in resiprocal space. Collective diffusion and viscosity are analyzed with in-house codes.

The results show that the time scales for different dynamic variables are not the same, but comparable. Compared with SPC the observed speed up factors for the Martini water are 2, 4 and 4 for shear viscosity, self-diffusion and collective diffusion, respectively. For polarizable Martini the speed up factors are 1.2, 5 and 6 in the same order as above. The computation method for collective diffusion is sensitive to changes in parameters and thus care is needed in interpeting the results. The results support the view of the semi-quantitative nature of the Martini model. The speed up factor of four is reasonably consistent and in simulations it can be used as a first approximation.

PREFACE

This work has been conducted as part of the research in Soft Matter and Biological Physics–research group at Tampere University of Technology. The topic of coarse-grained Molecular Dynamics has a central role group's research profile. I wish to thank professor Ilpo Vattulainen for giving the opportunity to work with this challenging task. I chose this topic because I wanted to delve into the physical aspect of biophysics — and that's exactly what I got. The theoretical and technical elements of this work have constantly stretched my abilities.

I wish to thank Bernard Reischl for support and insightful discussions during the time we shared the office, and Sanja Pöyry and Jill Steele for support and useful comments. A special thanks goes also to Monika Schatz for inciting the finalizing touch on this thesis. I'm especially thankful for the support from my mother. The path has been long and tedious, with many interruptions, but finally the project is finished.

Eero Hytönen
Opiskelijankatu 4 B 208
33720 Tampere
+358 50 467 8185

CONTENTS

TERMS AND SYMBOLS

A	Helmholtz free energy
$\mathcal{A}$	Amplitude of the force
A,B	Dynamic variables
ACF_A	Autocorrelation function of dynamic variable A
amu	Atom mass unit
B	Density-fluctuation autocorrelation function
β	Inverse temperature $(k_B T)^{-1}$
c_{ff}	Speed up coefficient associated with force field
c_m	Speed up coefficient associated with coarse-grained mapping
C_T	General transport coefficient
C_v	Heat capacity
CFD	Computational Fluid Dynamics
CG	Coarse Graining
CSC	Center for Scientific Computing
d	Number of dimensions
D_c	Collective diffusion coefficient
$\mathbf{D}_c$	Collective diffusion tensor
D_c^{CG}	Collective diffusion coefficient of a coarse-grained particle
D_s	Self-diffusion coefficient
D_s^{CG}	Self-diffusion coefficient of a coarse-grained particle
$D(\rho)$	Density-dependent diffusion
$\Delta\mathbf{R}$	Distance traveled by a particle
ΔG_{part}	Partition free energy
$\delta\rho$	Density fluctuation
δ_{ij}	Kronecker delta
$\delta(t)$	Delta function
e	Elementary charge
E_{kin}	Kinetic energy
E_{kin}^0	Target kinetic energy
$\hat{e}_x$	Unit vector in x-direction
ϵ	Depth of the interaction in Lennard-Jones potential
ϵ	Particle number fluctuation
ϵ_{LJ}	Interaction strength of the Lennard-Jones potential
ϵ_0	Dielectric constant
ϵ_r	Global dielectric constant

$\mathrm{erfc}(x)$	Complementary error function
ν	Kinematic viscosity
η	Shear viscosity
$f_0^{(N)}$	Equilibrium probability density
$F_{AB}(t,t')$	Equilibrium time-correlation function
$\mathbf{F}_i$	Force acting on particle i
$\mathbf{F}_R(t)$	Random force
FFT	Fast Fourier Transform
FG	Fine-grained
γ	Shear rate
GROMACS	Groeningen Machine for Computer Simulations
$H(\mathbf{p},\mathbf{q})$	Hamiltonian
$\mathbf{J}$	Total particle flux
$\mathbf{j_r}$	Particle current density at $\mathbf{r}$
K	Force constant
k	Wave vector
K_{angle}	Force constant for bond angle
K_{bond}	Force constant for bonded interactions
k_B	Boltzmann's constant
κ_{ij}	Isothermal compressibility
κ_T	Isothermal compressibility
L	Length of a particle
l_z	Height of simulation box
LJ	Lennard-Jones
m	Mass
MC	Monte Carlo
MD	Molecular Dynamics
$\measuredangle HOH$	Angle between hydrogens and oxygen in water
μ	Dynamic viscosity
N	Number of particles
N_{df}	Number of degrees of freedom
NVE	Constant number of particles, volume and energy
NVT	Constant number of particles, volume and temperature
p	Isotropic pressure
$\mathbf{p}$	Momentum vector

$\mathbf{p}_i$	Momentum vector for particle i
p_{xy}	Average of the pressure tensor estimator
P_{zx}	The shearing zx-component of the pressure tensor
P_{zz}	The zz-component of the pressure tensor
$\Phi(r_i j)$	Buckingham potential between particles I and j
PME	Particle Mesh Ewald
PP-method	Periodic Perturbations method
q_H	Partial charge of hydrogen
q_i	Charge of particle i
q_O	Partial charge of oxygen
QM	Quantum mechanics
$\mathbf{r}$	Coordinate vector
R	Radius of a particle
r^0_{bond}	Equilibrium distance for chemical bonds
$\mathbf{R}_{CG}$	Distance traveled by a coarse-grained particle
r_{cut}	Cut-off distance of interactions
r_{shift}	Distance at which potential is shifted to zero.
R_0	Radius of a particle
r_{ij}	Distance between particles i and j
r^0_{ij}	Equilibrium distance between particles i and j
R^{CG}_j	Distance traveled by a coarse-grained particle j
RDF	Radial distribution function
Re	Reynolds number
ρ	Density
$\rho(\mathbf{r})$	Density at $\mathbf{r}$
S	Dynamic structure factor
s	Shearing stress
s_{max}	Maximum shear rate
Sc	Schmidt number
σ	Effective size of particle in Lennard-Jones potential
σ_η	Error in viscosity
SPC	Single point charge
T	Temperature
t	Time
T_s	Temperature shift in PP-method

τ	Time constant
τ_p	Pressure coupling time constant
τ_r	Relaxation time
τ_T	Thermostat coupling time constant
TCAF	Transverse-current autocorrelation function
θ_{ijk}	Angle between particles i, j and k
θ^0_{ijk}	Equilibrium angle between particles i, j and k
U	Potential
$\mathbf{u}$	Velocity field
$U_{mod}(r)$	Modified Coulomb potential
$U_c(r_{ij})$	Coulomb potential between particles i and j
U_e	Electrostatic energy sum
$U_{LJ}(r_{ij})$	Lennard-Jones potential between particles i and j
v	Scalar velocity of a particle
V	Volume
$V_{angle}(\theta_{ijk})$	Potential for bond angle
V_{bond}	Potential for bonded interactions
v_{avg}	Average velocity of particles
ξ	Friction coefficient
Ξ	Virial tensor
ζ	Weight parameter

1. INTRODUCTION

During the last decade computational research has established itself as the third cornerstone of scientific research in addition to experiments and theory [1]. Computational studies are often building the bridge between the two and in fields such as economics, climate research, and biological sciences they are an indispensable tool in testing hypotheses, because the phenomena at hand are often too big, too small or too complex to study through any other means. Interestingly, the fields of science that use a lot of computing resources are also known for intensive experimental research. At CSC, the Finnish IT-center for science, the top three sciences using their computing resources in 2011 were nanoscience (38 %), physics (35 %) and chemistry (8 %) — and a good deal of physics can be assigned to biological physics [2]. This shows how computational science complements experiments and theory in opening new directions in research.

The exponential increase in computing power makes it possible to numerically solve problems which before were unimaginable. This further expands the areas of application for scientific computing. One such example is computational fluid dynamics (CFD), where complex fluid flows are analysed. Fluid flows are considered one of the most difficult problems to solve, but CFD gives tools to analyse the flows bringing forth advances in automobile design, aeronautics, and power plant furnace design. On the other hand as the computing power increases so does the complexity of the problems of interest, which creates the need for simple but flexible solutions that make it feasible to study big problems without an excessively large computational cost. To give an example, in soft matter computer simulations with Molecular Dynamics (see Chapter 3), the length- and time-scales have increased tremendously: today simulations of systems lasting up to a millisecond and containing up to million particles [3, 4] are readily feasible whereas 30 years ago a typical simulation might have contained no more than a few hundred particles and lasted for tens of picoseconds [5] – that is a combined increase of nine orders of magnitude. This would not have been possible without great advances in the molecular models and in computing power.

For the interest of this thesis, we turn to biophysical simulations where the development depicted above has lead to asking new kinds of questions. What kinds of simplified models can be used to reduce the computational load while maintaining

an acceptable level of accuracy? Also, what kinds of new physical aspects will the larger time- and length-scales bring along?

The answer to the first question lies in reducing the level of detail in the model, which is called coarse graining. Coarse graining schemes reduce the number of interacting particles by treating groups of atoms as a single particle and simplify the interactions between the particles [6]. The most common molecule in many soft matter simulations is water as solvent and hence many of the coarse graining methods focus on reducing the computational load associated with the solvent.

The coarse graining approaches can be divided into two groups: implicit solvent models, where the effect of solvent is embedded into non-solvent molecules, and explicit solvent models, where actual solvent molecules are present in the simulation [7]. By definition the implicit coarse graining schemes cannot include hydrodynamic effects and need special arrangements in order to get rid of the solvent. These special arrangements also create different kinds of limitations to the use of the models which do not burden explicit solvent models. The explicit solvent models in turn typically define two to three types of interaction sites: hydrophobic, hydrophilic and water-like. In some models these basic types contain subtypes which take into account the chemical nature of the underlying atoms. Interestingly, despite the variety of coarse graining approaches, all of the models produce fairly similar results in structural properties being largely consistent with all atom models [7].

The second question about new physical aspects is much more open to debate. One (side-)effect of the coarse graining process is that it makes the interactions smoother which results in accelerated dynamics in the system [6]. The molecules simply diffuse faster resulting in faster time evolution and problems with interpreting the time scale of the simulation. The speed up factor is typically between four and ten [6, 7]. An all atom (or fine-grained) simulation has typically a small simulation box, which prevents the emergence of large scale hydrodynamical phenomena, whereas these kinds of dynamic properties become increasingly important in coarse grained simulations. Even though coarse grained models capture the structural properties relatively well, how is the case for dynamic properties such as shear viscosity, self-diffusion or collective diffusion? The transfer of mass and momentum are important from physical point of view, but these have not been examined in detail.

In this thesis, the focus is on the dynamic properties of liquids and in the most ubiquitous molecule in living systems: water. The significance of water in biophysical simulations is illustrated by the fact that water typically constitutes more than half of the molecules in a simulation. Dynamic properties of the models become more important, because the time and size scales make it possible for dynamic and collective properties to have an impact. The standards used to characterize the models need to include the dynamic properties to a greater extent. To elucidate

these points the main research questions are

1. How do the dynamic properties of coarse grained water models correspond to real water and on the other hand to atomistic water models?

2. Are there differences in the time scales of different dynamic properties?

The following chapters are divided as follows. In the next chapter the physical concepts regarding dynamics and correlation functions are presented including the derivations of the necessary equations. Chapter 3 focuses on molecular models, Chapter 4 descirbes the simulation methods in this thesis and the simulations used, and Chapter 5 analyses and discusses the results. The final and sixth chapter contains the conclusions.

2. BROWNIAN MOTION AND TRANSPORT PHENOMENA

It is a random world down at the molecular level: the molecules in biological organisms are in constant motion. In soft matter systems the thermal energy makes the molecules wander around in a random walk and it is a small miracle how these random motions actually can create ordered structures — and even life. In order to understand the dynamics of soft matter systems, one has to understand Brownian motion, the random walk of particles, and how to analyze it. Brownian motion is the most fundamental mechanism of molecular transport, and the origin of self-diffusion as well as collective diffusion lies in Brownian motion. Collective diffusion and self-diffusion are the passive molecular mechanisms of transporting mass.

When discussing dynamics, the other important quantity to consider, in addition to spatial movements, is momentum. The intermolecular interactions and collisions create viscous friction, which drags down the flow of molecules. This is called viscosity of the liquid and it is a measure of the dissipation of momentum in the system. Shear viscosity and diffusion are called transport coefficients because they describe the transportation of a given quantity over time in this case the momentum and mass, respectively. Each liquid has its own characteristic values for both transport coefficients.

Next, starting from Brownian motion, the diffusion process is explained and a connection to viscosity is made. This is followed by a look on methods of analyzing diffusion and related processes. Diffusion and other transport phenomena are analyzed using correlation functions and the Green-Kubo –formalism.

2.1 Brownian Motion and Diffusion

Particles in a liquid are in constant motion randomly colliding with each other. It is the available thermal energy that facilitates this constant motion, and the magnitude of the motion depends on the temperature of the system according to the equipartition theorem $\langle v_{avg}^2 \rangle = 3k_BT/2$, where $\langle v_{avg}^2 \rangle$ is the average particle velocity, k_B is the Boltzmann's constant and T the temperature of the system. The velocity distribution of the particles is described by the Maxwell-Boltzmann distribution.

If the motion of a single particle were to be tracked, it would constitute a random walk, where the direction of the next steps could not be predicted from the previous

steps. The random collisions of a particle with its neighbors leads to a random trajectory. This also means that the trajectory has no memory effects. Such a process is also called a Markovian or a stochastic process. A classic analogy of a random walk is the walk of a drunken person on a square: sure, there is constant motion, but whether there is actual progress in certain direction will be an open-ended question. There are several fields, where random walks are encountered ranging from economics to ecology and engineering.

Single Particle Motions

The motion of a diffusive Brownian particle in a diluted suspension can be modeled with the phenomenological Langevin equation

$$m\frac{d^2\mathbf{r}}{dt^2} = -\xi m\frac{d\mathbf{r}}{dt} + \mathbf{F}_R(t), \tag{2.1}$$

where ξ is the friction coefficient, $\mathbf{r}$ is the location of the particle, and $\mathbf{F}_R(t)$ is a random force. The Langevin equation assumes that the diffusing particle is much larger than its neighbors. The first term on the right hand side is a friction term, which reflects the drag from the uniform medium as the Brownian particle moves through it. The second term on the right hand side introduces a random force which models the collisions of the particle with the surrounding molecules. The random force is considered as 'white noise' meaning that it is uncorrelated and vanishes in the mean:

$$\langle \mathbf{F}_R(t)\rangle = 0 \tag{2.2}$$

$$\langle \mathbf{F}_R(t+s)\cdot\mathbf{F}_R(t)\rangle = 2\pi R_0\delta(s) \tag{2.3}$$

where R_0 is the radius of the Brownian particle, $\delta(s)$ is the delta function, and the brackets denote ensemble average. Also, the random force is uncorrelated with the velocity

$$\left\langle \mathbf{F}_R(t)\cdot\frac{d\mathbf{r}}{dt}\right\rangle = 0. \tag{2.4}$$

The two terms on the right hand side of Eq. 2.1 reflect two different time scales in the system: the first term depicts the long term continuous drag inflicted on the particle, while the latter term models the instantaneous collisions from the surrounding particles, which also reflects the local structure of the liquid. These terms, however, are not independent. When solving the Langevin equation we see that

$$\xi = \frac{\beta}{3m}\int_0^\infty \langle \mathbf{F}_R(t)\cdot\mathbf{F}_R(0)\rangle dt, \tag{2.5}$$

where $\beta = (k_B T)^{-1}$ is the inverse temperature. The friction coefficient depends on the cumulative effect of the random collisions. On closer thought this is not an unexpected result, since the origin of both terms is the same — random collisions from neighboring molecules. When the Brownian particle travels through the medium the random collisions induce a systematic drag force, which has to be proportional to the frequency of the collisions. This is an example of the fluctuation-dissipation theorem [8].

What we are after here is not the friction coefficient, but a quantity which describes the movement of the Brownian particle. This is called the self-diffusion coefficient and it can be discovered using the Langevin approach (Eq. 2.1) and the Einstein relation (Eq. 2.9). Locating the particle initially ($t = 0$) at $\mathbf{r} = 0$ and using the conditions set to the force above, we find the mean square displacement [9]

$$\langle |\mathbf{r}|^2 \rangle = \frac{6k_B T}{\xi m} \left(t - \frac{1}{\xi} + \frac{1}{\xi} e^{-\xi t} \right). \tag{2.6}$$

At large times, when $\xi t \gg 1$, the above equation reduces to

$$\langle |\mathbf{r}|^2 \rangle = \frac{6k_B T}{\xi m} t. \tag{2.7}$$

Einstein determined the self-diffusion coefficient in the long time limit to be

$$D_s = \lim_{t \to \infty} \frac{1}{2dt} \langle |\mathbf{r}(t)|^2 \rangle, \tag{2.8}$$

where d is the number of dimensions in the system. Comparison between Eqs 2.7 and 2.8 yields the so called *Einstein relation*

$$D_s = \frac{k_B T}{\xi m}. \tag{2.9}$$

The significance of the Einstein relation is further expanded by the observations of Stokes, who discovered that an estimate for friction coefficient ξ for a spherical particle of radius R in a liquid with a shear viscosity η is given simply by

$$\xi = \frac{6\pi\eta R}{m}. \tag{2.10}$$

This equation is also known as *Stoke's law*. Substituting it into Eq. 2.9 yields

$$D_s \eta = \frac{k_B T}{6\pi R}. \tag{2.11}$$

This shows the connection between two important transport coefficients, which are both of interest in this thesis, namely the self-diffusion and the shear viscosity.

Intuitively it is easy to understand that the more viscous the fluid is the slower the particles move, and therefore it is expected to see the coefficients being inversely proportional. Raising the temperature increases thermal motion leading to faster diffusion and lower viscosity. It also enables the evaluation of the characteristic size of the molecules R which might be different from the physical radius determined in the model. Albeit Stoke's law is derived from macroscopic considerations and applies only to cases, where the Brownian particle is much larger than the surrounding molecules, it still gives a reasonable description on the molecular scale [9].

Collective Motions

The Langevin equation is used to describe the motion of a single particle, but what if we have several diffusive particles? How can many-particle diffusion be characterized? The relaxation of concentration gradients is an example of a many-particle diffusion phenomenon. Here the problem is tackled with hydrodynamics using Fick's law and mass transport equation. The hydrodynamic regime means that the time and length scales involved are much larger than the ones experienced by individual molecules. Starting from the fact that mass is a conserved quantity, we write the continuity equation for mass transport

$$\frac{d\rho(\mathbf{r})}{dt} + \nabla \cdot \mathbf{j}_\mathbf{r} = 0, \tag{2.12}$$

where $\rho(\mathbf{r})$ is the density at $\mathbf{r}$ and $\mathbf{j}_\mathbf{r}$ is the particle current density at the same place $\mathbf{r}$. Here, the density ρ represents the mass and its rate of change depends on the gradient per unit area. The continuity equation simply states that the density in a unit area changes at a rate which corresponds to the net flow of particles in (or from) that same area. The particle current density can be expressed as the gradient of density with (local) diffusion constant according to *Fick's first law*

$$\mathbf{j}_\mathbf{r} = -D(\rho)\nabla\rho(\mathbf{r},t), \tag{2.13}$$

where $D(\rho)$ is the density-dependent diffusion. Combining the above equations yields *Fick's second law*

$$\frac{\partial\rho}{\partial t} = -\nabla \cdot D(\rho)\nabla\rho(\mathbf{r},t). \tag{2.14}$$

In the most general case Eq. 2.1 is a non-linear partial differential equation, which often can not be solved analytically. Thus, a further assumption is made that the diffusion coefficient does not depend on the density of the particles (ie. $D(\rho) = D$).

Now the equation simplifies to

$$\frac{\partial \rho}{\partial t} = -D\nabla^2 \rho(\mathbf{r}, t). \tag{2.15}$$

We have reached another definition for the diffusion coefficient: it is the rate of change coefficient for density fluctuations. Now, however, we are dealing with a different kind of diffusion coefficient than what was encountered with Langevin equation. All the quantities originate from hydrodynamics and implicitly represent macroscopic length and time scales — not microscopic as with self-diffusion. The diffusion factor obtained here shall be referred to as the *collective diffusion coefficient* and denoted D_c later in the thesis.

In the derivation of Fick's second law two assumptions have been made. First, the use of hydrodynamics assumes that the microscopic details can be neglected, and second, the continuity equation (Eq. 2.12) assumes that mass transport is proportional to the gradient of the density and higher than second order terms of the density gradients can be neglected. This is also one expression of the linear response theorem, which states that the system responds to an external perturbation in the same way as to a fluctuation of the same size. In hydrodynamics, the drive towards equilibrium means that spontaneous fluctuations in the microscopic variables decay in the mean at sufficiently large length scales and long time scales according to laws governing the macroscopic variables [10].

Both descriptions for self-diffusion and collective diffusion rely on phenomenological equations — that is, they only provide mathematical descriptions of the observed results without considering the fundamentals behind the equations — and are not directly applicable as such to calculating diffusion coefficients from computer simulations (with the exception of Eq. 2.8). A simulation produces the trajectories of individual molecules and not the variables found in the above equations. Means of connecting the macroscopic and microscopic levels are needed in order to have a practical way to compute the transport coefficients. This is achieved with the Green-Kubo –formalism [9] discussed later in this chapter.

Viscosity

Viscosity is a term we do not encounter too often in our daily lives, but we are all familiar with viscosity even though we might not think about it. Take liquids for example: we know syrup is much more difficult to mix than water. And we all know how margarine becomes more pliable when we leave margarine on the table for a while — the viscosity is decreased. Despite the intuitive feel, getting a grip of the physical formulation of viscosity is much more difficult. Consider the big difference between solids and liquids in their response to deformations. Solids want to relax

back to their equilibrium position after a deformation, but liquids do not have this property. Instead they want to maintain their current state of motion. Liquids have no memory effects at variance to solids. That is, liquids do not remember where they were, only where they are going. Viscosity relates to the propensity at which the liquids want to maintain the current state of motion. Liquids respond to the *rate* of deformation where as solids respond to the *amount* of deformation.

Another interesting feature of liquids is that for different-size objects the same medium can feel very different to move in. For a small amoeba water is a difficult medium to move in, but humans swim in water with ease. The reason for this is that the ratio of frictional forces — caused by molecular interactions — to momentum of the object changes with the length scale. This property is captured in Reynolds number

$$Re = \frac{\rho v L}{\mu}, \tag{2.16}$$

where ρ is the density of the fluid, v is the mean velocity of the object relative to the fluid, L is the characteristic length scale and μ is the dynamic viscosity of the fluid. The numerator in the above equation describes the total momentum transfer and the denominator the molecular momentum transfer. Reynolds numbers range from 10^{-1} to 10^{9}. Reynolds number helps to understand the role of viscosity and the interplay between inertia and viscous forces in liquid systems. When the Reynolds number is large (on the order 10^4) momentum is the dominating property and objects move quite easily in the medium in question. With low Reynolds numbers the frictional properties take the upper hand making it difficult for particles to move in the medium.

The viscosity in Eq. 2.11 seems to play a balancing or opposite role to self-diffusion, since they are inversely proportional. Self-diffusion measures the motion of the molecules, while viscosity is a measure for the frictional forces which oppose the motions of molecules. Following Tuckerman [11], let us derive an equation for shear viscosity by starting with a set of microscopic equations of motion

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m_i} + \gamma y_i \hat{\mathbf{e}_\mathbf{x}} \tag{2.17}$$

$$\frac{d\mathbf{p}_i}{dt} = \mathbf{F}_i - \gamma p_{yi} \hat{\mathbf{e}_\mathbf{x}}, \tag{2.18}$$

where $\mathbf{r}_i$, $\mathbf{p}_i$, m_i are the location, momentum and mass of particle i, y_i and p_{yi} are the y-components of the location and momentum of particle i, $\mathbf{F}_i$ is the force acting on particle i, and $\hat{\mathbf{e}_\mathbf{x}}$ is the unit vector in x-direction. The parameter γ is the shear rate. The equations describe a situation where liquid is placed between two plates which are pulled apart in opposite directions creating a *shear force*. The situation is illustrated in Fig. 2.1. In the bottom the liquid is at rest ($v = 0$). The velocity of

the 'liquid layers' increases in x-direction making them slide past each other, which results in a *shear stress* between the layers. The shear flow increases the momentum of the particles in the system and the total momentum is $\mathbf{p}_i + m_i\mathbf{v}(y_i)$.

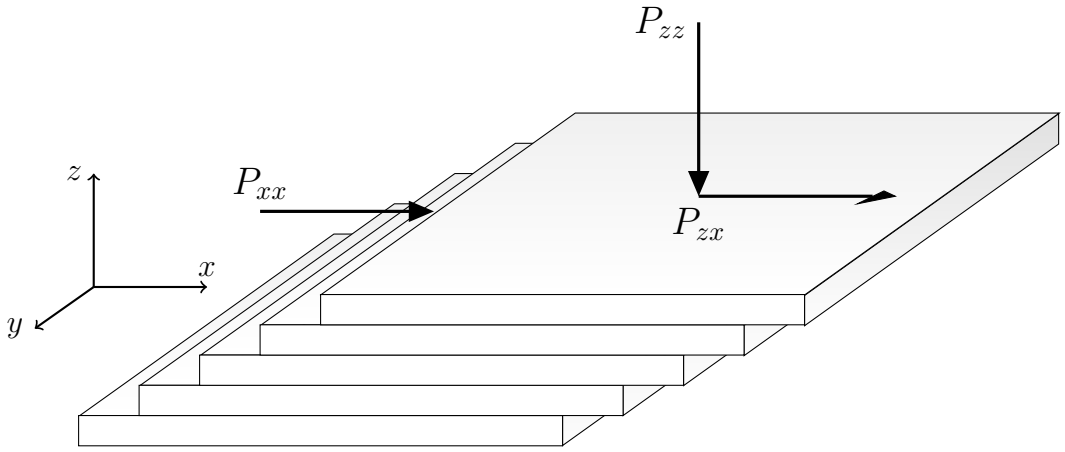


Figure 2.1: The shearing of liquid layers upon imposing a shearing force in x-direction. Liquid can be thought of as layers sliding past each other causing friction between the layers. The sliding resistance is proportional to the zx component of the pressure.

The shear flow generates anisotropy in the system, which further on creates asymmetry in the internal pressure. When the pressure is non-uniform a single scalar value for the pressure is insufficient and a new construct that takes into account the different pressures in different directions is needed. If the pressure P is obtained from an estimator for isotropic pressure, which stands

$$p = \frac{1}{3V}\sum_{i=1}^{N}\left(\frac{\mathbf{p}_i^2}{m_i} + \mathbf{r}_i \cdot \mathbf{F}_i\right), \tag{2.19}$$

where V is the volume of the system and in equilibrium $P = \langle p \rangle$, one can define in analogy an estimator for the pressure tensor (also called the stress tensor) with

$$P_{\alpha\beta} = \frac{1}{V}\sum_{i=1}^{N}\left(\frac{(\mathbf{p}_i \cdot \hat{\mathbf{e}}_\alpha)(\mathbf{p}_i \cdot \hat{\mathbf{e}}_\beta)}{m_i} + (\mathbf{r}_i \cdot \hat{\mathbf{e}}_\alpha)(\mathbf{F}_i \cdot \hat{\mathbf{e}}_\beta)\right), \tag{2.20}$$

where $\hat{\mathbf{e}}_\beta$ and $\hat{\mathbf{e}}_\alpha$ are the unit vectors in direction $\alpha, \beta = x, y, z$. The first subscript tells the normal of the plane on which the stress is acting on and the second subscript the direction of the pressure. Components, where $\alpha = \beta$, are normal components which constitute the isotropic pressure, and components with $\alpha \neq \beta$ are the shearing pressures or stresses.

The mathematical description may look complicated, but intuitively the components of the pressure tensor are easy to comprehend. Imagine a box on a surface. There are two different ways of pushing the box around. First, one can push it from

one of the sides and apply pressure perpendicular to the side and parallel to the surface, where the box lies. This corresponds to the diagonal terms xx and yy. The other option is to apply pressure in a oblique angle on the box's roof and try to shear it. The applied pressure is perpendicular to the surface, but the movement of the box happens parallel to the surface. Now the cross terms come into play. This corresponds to the cross terms zy and zx. These concepts are illustrated also in Fig. 2.1. It has been observed that the xy-cross term of the pressure tensor depends on the velocity field or shear rate

$$P_{xy} = -\eta \frac{\partial v_x}{\partial y} = -\eta\gamma, \tag{2.21}$$

where η is the shear viscosity. Solving for η yields

$$\eta = -\frac{P_{xy}}{\gamma} = -\lim_{t\to\infty} \frac{\langle p_{xy}(t)\rangle}{\gamma}, \tag{2.22}$$

where $\langle p_{xy}(t)\rangle$ is the average of the xy-component of the pressure tensor estimator in the above described steady-state flow case.

In order to calculate $\langle p_{xy}(t)\rangle$, linear response theory and the concept of dissipative flux are applied [11]. Now the dissipative flux becomes

$$j(x) = \sum_{i=1}^{N} \left[\gamma y_i (\mathbf{F}_i \cdot \hat{\mathbf{e}_\mathbf{x}}) + \gamma p_{yi} \frac{\mathbf{p}_i \cdot \hat{\mathbf{e}_\mathbf{x}}}{m_i} \right] \tag{2.23}$$

$$= \gamma \sum_{i=1}^{N} \left[\frac{(\mathbf{p}_i \cdot \hat{\mathbf{e}_\mathbf{y}})(\mathbf{p}_i \cdot \hat{\mathbf{e}_\mathbf{x}})}{m_i} + (\mathbf{r}_i \cdot \hat{\mathbf{e}_\mathbf{y}})(\mathbf{F}_i \cdot \hat{\mathbf{e}_\mathbf{x}}) \right] \tag{2.24}$$

$$= \gamma V p_{xy}. \tag{2.25}$$

With the so called linear response formula [11] the average of the xy-component of the pressure tensor is

$$\langle p_{xy}(t)\rangle = \langle p_{xy}\rangle_0 - \frac{\gamma V}{k_B T} \int_0^t \langle p_{xy}(0) p_{xy}(t-s)\rangle_0 ds, \tag{2.26}$$

where V is the volume of the system and $\langle p_{xy}\rangle_0$ is the equilibrium average of the pressure tensor, which is zero because in equilibrium the pressure is symmetric. Finally, substituting into Eq. 2.21 and taking the limit, the result for shear viscosity is

$$\eta = \frac{V}{k_B T} \int_0^\infty \langle p_{xy}(0) p_{xy}(t)\rangle_0 dt. \tag{2.27}$$

The result is an example of a Green-Kubo equation, which will be met more closely in the next section. This form manages to show the relation between the microscopic phenomenon (the pressure tensor) and a macroscopic variable (shear

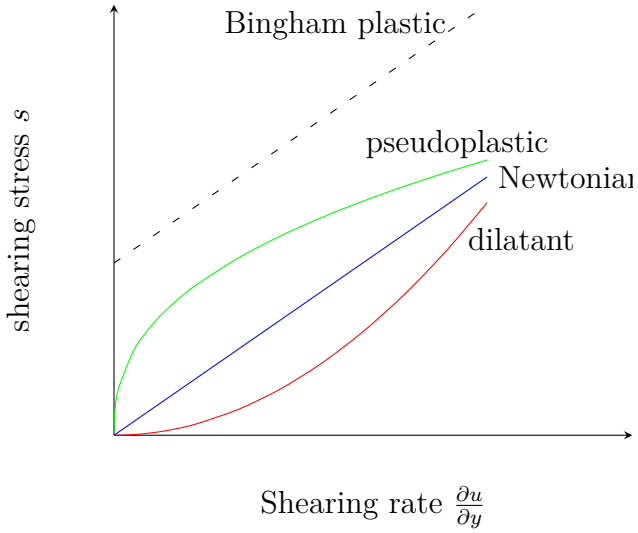


Figure 2.2: Response to shearing stress of different liquids: Newtonian (blue), dilatant (red), pseudoplastic (green), Bingham plastic (dashed). Dilatant liquids become thicker upon stirring and the shearing stress increases. Inverse happens to pseudoplastics. Bingham plastics are rigid below certain threshold but become viscous fluids at high shearing rates.

viscosity) and hence describes completely the phenomenon at hand.

Viscosity is a property specific to the liquid and the molecular structure of the liquid determines the precise functional form of the response to shear force. For Newtonian fluids, which are the most common, this ratio is linear as expressed in Eq. 2.21. In contrast non-Newtonian fluids have a shear viscosity that can be a function of time or shear stress. Fig. 2.2 illustrates the different responses as a function of shear rate. The viscosity of *shear-thinning* fluids or pseudoplastics decreases with increased shear stress. They behave as solids when undisturbed, but under stirring or other agitation turn into fluids. The reverse happens to *shear-thickening* or dilatant fluids.

We already have a measure how mass spreads around in the form of self-diffusion coefficient and it would be useful to have a corresponding measure for momentum. Dividing shear viscosity by liquid density does just that and yields a measure for the diffusion of momentum, called *kinematic viscosity*

$$\nu = \frac{\eta}{\rho}. \tag{2.28}$$

Kinematic viscosity characterizes liquids in a similar fashion to Reynolds number (Eq. 2.16) by considering the ratio of viscous force to inertia. Further on, combining kinematic viscosity and diffusion provides another way to characterize fluid flows and

even distinguish between phases. This is called the Schmidt number and it is defined

$$Sc = \frac{\nu}{D_s}. \tag{2.29}$$

The Schmidt number varies a great deal depending on the medium since the viscosity of fluids spans three orders of magnitude. For gases the Schmidt number is again many orders of magnitude smaller due to low viscosity and high particle velocities. By observing the Schmidt number it is possible to determine how freely the fluid flows.

2.2 Definitions of Transport Coefficients with Green-Kubo Equations

The time-dependent transport phenomena are described by so called transport coefficients, such as diffusion or viscosity coefficients. Transport coefficients tell how changes in one place are propagated through the system. Whether it is heat or mass transfer, there is a transport coefficient related to it. In mathematical terms this can be expressed with time-correlation functions and the Green-Kubo equation. The Green-Kubo equation connects the microscopic level of statistical mechanics with the macroscopic level measured in experiments.

Let us start with the time-correlation function, which describes how the value of a time-dependent (i.e. dynamic) variable A at some time t is related to the value of a variable B at a later time t'. In classical statistical mechanics the equilibrium time-correlation function $F_{AB}(t,t')$ of dynamic variables A and B is defined as

$$F_{AB}(t,t') = \langle A(t)B(t')\rangle \tag{2.30}$$

where the angular brackets denote ensemble average

$$\begin{aligned}\langle A(t')B(t'')\rangle &= \iint A[\mathbf{r}^N(t'),\mathbf{p}^N(t')]B[\mathbf{r}^N(t''),\mathbf{p}^N(t'')] \\ &\quad \cdot f_0^{(N)}[\mathbf{r}^N(t''),\mathbf{p}^N(t'')]d\mathbf{r}^N(t'')d\mathbf{p}^N(t'')\end{aligned} \tag{2.31}$$

or time average

$$\langle A(t')B(t'')\rangle = \lim_{\tau\to\infty}\frac{1}{\tau}\int_0^\tau A(t'+t)B(t''+t)dt. \tag{2.32}$$

In the above, $f_0^{(N)}$ is the equilibrium probability density and the ensemble average is carried out over all possible states in the phase space ($\mathbf{r}^N$ and $\mathbf{p}^N$) at time t''. If A and B are the same variable the time-correlation function is called the autocorrelation function or self-correlation function and denoted ACF_A. In the case of diffusion

we will be dealing with autocorrelation functions. With dynamic variables we can define the associated transport coefficient in the Green-Kubo form

$$C_T = \alpha \int_0^\infty \langle A(0)A(t)\rangle dt. \tag{2.33}$$

where α is the appropriate thermodynamic factor, which depends on the specific transport coefficient in question.

Self-diffusion

The motions of a single molecule can be characterized with the self-diffusion coefficient, which in turn can be calculated with the Einstein equation (Eq. 2.8) or with the Green-Kubo formalism using the velocity autocorrelation function

$$D_s = \frac{1}{dN} \sum_{i=1}^{N} \int_0^\infty \langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0)\rangle dt, \tag{2.34}$$

where N is the number of particles, d is the number of dimensions of the system and $\mathbf{v}_i(t)$ is the velocity of particle i at time t. The self-diffusion coefficient in the Green-Kubo equation is interpreted as the time integral of velocity autocorrelation and averaged over all the particles. The correspondence between the Einstein form and the Green-Kubo form is a general property of correlation functions. In fact, it can be shown that the two are exactly equivalent [9]. If $A(t)$ is a dynamic variable and $\dot{A}(t)$ its time derivative then the following holds [12]

$$\lim_{t\to\infty} [A(t) - A(0)]^2 = \int_0^\infty \langle \dot{A}(t)\dot{A}(0)\rangle dt \tag{2.35}$$

The Green-Kubo equation captures the same idea as the Einstein form since the velocity integral over time yields distance and the correlation of velocity over time reflects how far the particle travels.

Collective Diffusion

As discussed previously, the collective diffusion coefficient gives insight into the concerted motions of particles. Therefore the dynamic quantity involved in the calculation must include the motions of all the particles. The Green-Kubo equation in an isotropic system is [13]

$$D_c = \frac{1}{dk_B T \rho N \kappa_T} \int_0^\infty \langle \mathbf{J}(t) \cdot \mathbf{J}(0)\rangle dt, \tag{2.36}$$

where d is the dimensionality of the system, ρ is the number density, N is the number of particles in the system, κ_T is the isothermal compressibility and $\mathbf{J}$ is the

total particle flux defined as $\mathbf{J} = \sum_{i=1}^{N} \mathbf{v}_i(t)$. In practice calculating D_c from this definition is computationally heavy and sensitive to error.

Collective diffusion can also be obtained from density fluctuations. A density fluctuation at point $\mathbf{r}$ at time t is the difference between the local density and the global average density $\delta\rho(\mathbf{r},t) = \rho(\mathbf{r},t) - \langle\rho\rangle$. Using this, the density-fluctuation autocorrelation function is defined as [14]

$$B(\mathbf{r}-\mathbf{r}',t) = \langle\delta\rho(\mathbf{r},t)\delta\rho(\mathbf{r}',0)\rangle. \tag{2.37}$$

The density-fluctuation and its close relative, the van Hove autocorrelation function [15] are important in the study of structure and dynamics of materials, because taking the (temporal and spatial) Fourier transform of the van Hove function yields the dynamic structure factor $S(k,\omega)$, which can be measured experimentally with neutron scattering experiments. Thus the afore mentioned functions provide a bridge between experiments and theory.

In order to derive an equation for the collective diffusion coefficient let us write the Fick's second law for the density fluctuations

$$\frac{\partial\delta\rho(\mathbf{r},t)}{\partial t} = D_c\nabla^2\delta\rho(\mathbf{r},t). \tag{2.38}$$

Multiplying both sides with $\delta\rho(0,0)$ and taking the ensemble average we get

$$\frac{\partial\langle\delta\rho(\mathbf{r},t)\delta\rho(\mathbf{0},0)\rangle}{\partial t} = D_c\nabla^2\langle\delta\rho(\mathbf{r},t)\delta\rho(\mathbf{0},0)\rangle \tag{2.39}$$

$$\frac{\partial B(\mathbf{r},t)}{\partial t} = D_c\nabla^2 B(\mathbf{r},t). \tag{2.40}$$

Assuming hydrodynamic regime (small $\mathbf{k}$ and long times) the spatial Fourier transform of B decays with [14]

$$B(\mathbf{k},t) = B(\mathbf{k},0)e^{-\mathbf{k}^2 D_c t} \tag{2.41}$$

where $\mathbf{k}$ is the wave-vector. The collective diffusion factor is recovered from the exponential and now it is possible to measure D_c from the decay rate of density fluctuations in k-space.

2.3 The Relation Between Collective and Self-Diffusion

One might think that collective diffusion is just the sum of the individual one-particle diffusions and thus easily associated with self-diffusion. This picture is unfortunately far from the truth. Often the intermolecular interactions are complex and as a consequence ingredients factoring into collective motions are not readily

determined. Juxtaposing the Green-Kubo equations for collective and self-diffusion results in

$$D_c = D_s \epsilon \left[1 + \frac{\sum_{i \neq j} \int_0^\infty \langle \mathbf{v}_i(t) \cdot \mathbf{v}_j(0) \rangle dt}{\sum_{i=1}^N \int_0^\infty \langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0) \rangle dt} \right], \tag{2.42}$$

where ξ is the particle number fluctuation defined as

$$\epsilon = \frac{\langle N \rangle}{\langle N^2 \rangle - \langle N \rangle^2}. \tag{2.43}$$

We see that the role of the intermolecular interactions stands out and only if the cross-correlations $\langle \mathbf{v}_i(t) \cdot \mathbf{v}_j(0) \rangle$ are negligible, there is a simple relation between the diffusion coefficients. If the motions of the particles are completely uncorrelated the sum in the nominator goes to zero and we get the so called *Draken equation*

$$D_c = \xi D_s. \tag{2.44}$$

The Draken equation applies in the dilute limit (or ideal gas limit) or in high temperatures, where the motions become uncorrelated due to lack of interactions between the particles or large kinetic energies.

Now we have established the basic understanding of the origins of the dynamic variables and their relation on one hand, to each other, and on the other hand, to the fundamental physical variables of velocity, location and momentum. The mathematical formulations of Brownian motion and viscous friction lead to quantifying the ideas of mass and momentum transfer into transport coefficients. Knowing both the big picture and the microscopic origin is essential in order to have a solid stepping stone into following chapters. The equations derived in this chapter tells us what to look for in the molecular dynamics simulations and helps in understanding the analysis methods explained in Chapter 4. But before going into the simulations let us look at the molecular models in some detail.

3. MOLECULAR MODELS

The ultimate aim of any model is to represent the underlying phenomenon as realistically as possible. In practice, however, this is too demanding a task and models aim to describe some aspects of the phenomenon to a reasonable degree of accuracy often exchanging accuracy for simplicity. For water this is especially true: despite its seemingly simple structure, water is an extremely complex fluid to model. Next we will discuss some aspects of modeling water followed by overviews of the water models used in this thesis. Although the focus of the thesis is in dynamic properties of water, the first section is devoted to describing the water models from atomic and quantum mechanical levels to give an overview of the field and appreciate the work of the modeler.

3.1 Modeling Water

The development of accurate water models is an art in itself. A huge number of water models have been developed during the last 30 years at a rate of more than one per year, which displays how tedious task it is to model water [16]. There is no consensus on which is the best model and, surprisingly, some of the oldest models are still among the most used (SPC published in 1981, SPC/E from 1987, and TIP3P developed in 1983). The structure of water is extremely complex and dynamic, where quantum and many-body effects play an important role. Getting all the properties of water from heat of vaporization to rotational relaxation times and dielectric constant to match the experimental values in a single model is at the current level simply impossible.

Two schemes can be used in modeling water: *ab initio* approach or empirical approach. *Ab initio* or first principles approach relies on quantum mechanical (QM) computation of properties which are, at least in principle, exact. The potential function, which characterizes the interactions, is discovered as an analytical fit to the QM results. Unfortunately, QM calculations are extremely heavy and often they are restricted to a dimer or a tetramer of water molecules. This means that they are inherently precluded from determining the properties of bulk water. Therefore it is more fruitful to use QM only on a specific part of modeling such as charge transfer or hydrogen bonding. Empirical approach takes a more holistic view and the aim is to reproduce the electrostatic, structural, dynamical and thermodynamical properties

Table 3.1: Different types of water models. [16]

Number of sites	Type		
	Rigid	Flexible	Polarizable
3	SPC, SPC/E	TIP3P, SPC/F	SPC/FQ, POLARFLEX
4	TIP4P	MCDHO	MCDHO, SWFLEX
5	TIP5P		POL5, NEMO

of real water as closely as possible. When aiming for a multi-scale model, the approach of choice is empirical.

The water models can be categorized based on their structure and internal degrees of freedom. Since water molecule consists of three atoms, the natural choice is to have three interaction sites, but models with four or five interaction sites also exist. The bonds and bond angles can be rigid or flexible. Categories and examples for each are compiled to Table 3.1. Models are developed for a certain purpose, and have different strengths: some have exact thermodynamic properties while others can reproduce the hydrogen bonding very accurately. In general it can be said that the more detailed the model, the more properties can be reproduced. However, it is not clear that this is always the case. In many examples, adding flexibility or more interaction sites to the model does not improve the performance to a great extent, while having the offset of being computationally more expensive [16]. In simulations of biological systems the addition of an extra interaction site to the water model can increase the simulation time by half [17]. Simple rigid models, such as SPC, often are a good compromise between accuracy and efficiency, hence such a model is used in this work also.

Among the many properties of water the most relevant to match are [16]

- Density in the liquid state
- Heat of vaporization, which reflects the polarization of the model
- Diffusivity
- Structure of water in radial distribution function
- Dielectric constant.

Different relaxation times also offer ways to rate the performance of water models.

The water models have a lot in common when it comes to mathematical details of the model. The interactions between the atoms within a molecule and between molecules are created with interaction potential functions, which consider the intermolecular forces such as dispersion, multipole interactions, dipole-dipole interactions, hydrogen bonding and electrostatics. For the models containing flexibility and

internal degrees of freedom, potentials have to consider bond stretching and bond angle as well. The electrostatic potential is modeled with the Coulomb potential, while for the van der Waals forces there are a few options. The most common is the Lennard-Jones potential

$$U_{LJ}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] = \left(\frac{C_{12}}{r_{ij}}\right)^{12} - \left(\frac{C_6}{r_{ij}}\right)^{6}, \tag{3.1}$$

where ϵ is the depth of the interaction well and σ the distance at which the potential has a value zero, or alternatively they can be combined to give parameters C_6 and C_{12}. Another commonly used choice is the Buckingham potential

$$\Phi(r_{ij}) = Ae^{-Br_{ij}} - \frac{C}{r_{ij}^6}, \tag{3.2}$$

where A, B and C are adjustable parameters. The Lennard-Jones potential is computationally more efficient but less flexible. The most common choice for the intramolecular interactions (i.e. chemical bonds) is the harmonic potential of type

$$U(r_{ij}) = K(r_{ij} - r_{ij}^0)^2, \tag{3.3}$$

where K is the force constant, or spring constant in terms of classical mechanics, and r_{ij}^0 is the equilibrium value for the bond or for angle, depending on which is in question. Other types, such as anharmonic potentials or additional coupling terms between the bonds and angle, are also used.

The choice of which potentials to use is in the discretion of the modeler and depends on the application. The parameters of the model have to be set carefully, since the properties of water models are highly sensitive to the parametrization of the model and underlying assumptions. For example, increasing the O-H -bond length just by 2 % from 1.0 nm to 1.02 nm can reduce the self-diffusion coefficient by a third [17]. Now that a general overview has been presented we look at the models in this thesis in detail.

3.2 Single Point Charge

Single Point Charge, or SPC water model [5] is one of the oldest water models in use for molecular dynamics (MD) simulations. It was published in 1981 and since then many variants from it have been developed including SPC/E, with different charges assigned to oxygen and hydrogen, SCP/HW, and flexible models SPC/F, and SPC/Fw [17]. SPC was originally designed for simulations with hydrated proteins. The aim was to create a simple yet reliable water model for the liquid state, which can be extended to interactions with various other molecules. The simplis-

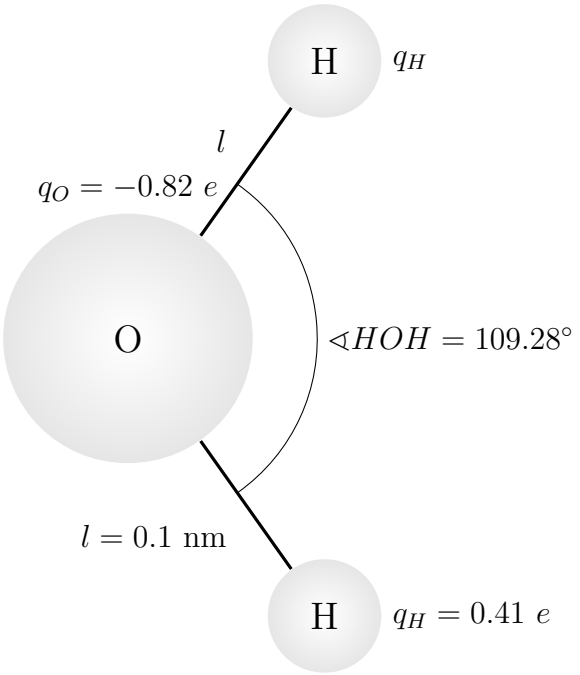


Figure 3.1: The structure of SPC water model

tic approach taken in the development of the model probably reflects the era when CPU-time was expensive and simplicity was important. Major number of the water models published roughly a decade later were computationally heavier flexible models [16].

SPC is a rigid, 3-site model, where the oxygen-hydrogen -bond length is fixed at 0.1 nm and the angle $\measuredangle HOH$ at 109.28°. The geometry is depicted in Fig 3.1. Note that the angle is larger than the experimentally measured 104.5°. The charges centered at the atomic nuclei carry a charge of $0.41e$ and -$0.82e$, for hydrogen and oxygen respectively. The model utilizes an effective pair potential of Lennard-Jones 6-12 form between oxygen centers, where the attractive part is derived from London expression and it has experimental origins. The repulsive part and the charge parameters were fitted using MD simulations to match the experimental interaction energy and pressure at 300 K. The full parameters and chosen properties of the SPC model are given in Table 3.2. We can clearly see that the dipole moment is well below that of real water as well as the dielectric constant, while the self-diffusion coefficient is significantly higher. This is by no means a coincidence, since the dielectric constant depends on the dipole moment, which again depends (among others) on the bond length. The bond length on the other hand is intimately related to the hydrogen bonding capabilities of the model. Hydrogen-bonding creates friction and reduces the translational motion of the molecules resulting in slower diffusion. The underestimated dielectric constant indicates indirectly that the hydrogen bonding is underestimated, which results in overestimated diffusion.

Table 3.2: The parameters for water and the SPC water model. Data from Refs [5, 17–22] Experimental values given where available at T = 298 K. Units in parenthesis.

Property	Unit	Value	Experimental
Partial charge of oxygen	(e)	-0.82	n/a
Partial charge of hydrogen	(e)	0.41	n/a
L-J parameter C6	(kJ mol^{-1} nm^{-6})	$2.6171 \cdot 10^{-3}$	n/a
L-J parameter C12	(kJ mol^{-1} nm^{-12})	$2.6331 \cdot 10^{-6}$	n/a
O-H bond length	(Å)	1.000	0.9584
H-O-H bond angle	(deg)	109.47	104.45
Density	(kg m^{-3})	977	997
Dipole moment	(Debye)	2.27	2.9
Self-diffusion coefficient at 298 K	(cm^2 s^{-1})	$4.02 \cdot 10^{-5}$	$2.3 \cdot 10^{-5}$
Dielectric constant at 298 K	ϵ_0	66.29	78.54
Enthalpy of vaporization	(kJ mol^{-1})	44.18	43.99
First peak distance in RDF	(nm)	0.276	0.288
Height of first peak in RDF		2.82	3.09

Indeed, the SPC model is very simple but powerful: many of the important properties of water are well reproduced with only a few parameters. The other side of the coin is that the model is not very flexible and thus there is not too much room for improvement. The internal degrees of freedom are frozen and the intermolecular interactions are modeled with the LJ-potential which lacks flexibility compared, for instance, to the Buckingham exponential-6 potential [16, 23]. The SPC model is non-polarizable and therefore it cannot respond to the changes induced by local environment electrostatics in heterogeneous systems. This is relevant because polarization effects are highly environment dependent. The density is off by 2 %, which is acceptable since the structure of water is relatively accurate as judged by the distance of the first peak in O-O radial distribution function r_1 and the height of the same peak $g(r_1)$ (see Table 3.2).

3.3 Martini Force Field

Martini force field is a coarse grained force field, which simplifies a molecule by averaging several atoms of the molecule into one particle and making the molecular model coarser. This results in computational efficiency at the cost of loss of detail.

Martini force field is based on the reproduction of partitioning free energies between polar and apolar phases of a large number of chemical compounds and it is designed for biomolecular simulations. Here the focus is on the parametrization, general characteristics of the Martini force field and aspects relevant to dynamic properties of the model. For a detailed description of the force field, see [24]. The

following is based on the original paper by Marrink *et al.* [24].

The Martini force field and its earlier version [6] have been used in over 40 studies with good results. It has been applied to systems ranging from vesicles and lipoproteins to membrane-protein assemblies. Central to many biologically relevant processes (such as peptide-membrane binding) is the degree of partitioning of the constituents between polar and apolar environments. Thus the accuracy of the partitioning coefficients is chosen as one of the criteria for the model. In coarse graining the entropy will be decreased by definition due to the loss of atomic degrees of freedom. In order to get correct free energies, a corresponding decrease in enthalpy has to be introduced to the coarse grained system.

The Martini force field has a number of chemical building blocks, which are used to construct the coarse grained molecules. The standard scenario is four-to-one mapping, that is four heavy atoms including the associated hydrogens are mapped to one coarse grained bead. There are four main bead types for charged, polar, apolar and non-polar atom groups. Each main type has 4–5 subtypes, which allows the fine tuning according to chemical characteristics. With the subtypes it is possible to control the polar affinity of the groups or the capability to form hydrogen bonds (as a donor, as an acceptor or as both). All beads (excluding beads connected by a chemical bond) interact with van der Waals forces modeled with the Lennard-Jones (LJ) potential

$$U_{LJ}(\mathbf{r}_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{3.4}$$

where σ_{ij} is the effective radius parameter set to 0.47 nm with one exception[1] and ϵ_{ij} is the strength of the interaction between particles i and j. The difference between the bead types is expressed in the ten Lennard-Jones -interaction levels between the beads. The interaction levels and their typical uses are listed in Table 3.3. The most polar interaction at $\epsilon = 5.6$ kJ mol^{-1} models compounds that are solid in the room temperature and charged groups with strong hydration shell. The second level is used to model bulk water making it noteworthy in the context of this thesis.

Non-bonded charged particles (or beads) have a full charge and they interact also via a shifted Coulomb potential

$$U_c(\mathbf{r}_{ij}) = \frac{1}{4\pi\epsilon_0\epsilon_r}\frac{q_i q_j}{r_{ij}}, \tag{3.5}$$

where $\mathbf{r}_{ij}$ is the distance between the particles, q_i and q_j are the charges of par-

[1] $\sigma = 0.62$ nm for interactions between charged and the most apolar types. The extended closest distance makes it more favorable for charged particles to keep their hydration shells when dragged into apolar medium.

Table 3.3: The interaction levels of LJ-potential in Martini force field and their typical uses as given in [24]

Interaction level	Interaction strength (kJ mol⁻1)	Use
O	5.6	Solids
I	5.0	Bulk water
II–III	4.5–4.0	Volatile liquids
IV	3.5	Non-polar aliphatic chains
V–VIII	3.1–2.0	Hydrophobic repulsions between polar/non-polar phases
IX	0.2	Between charged particles and very apolar medium

ticles i and j, respectively, ϵ_0 is the permittivity of vacuum and $\epsilon_r = 15$ is the relative dielectric constant. Both non-bonded interactions have a cut-off distance at $r_{cut} = 1.2$ nm. The Coulomb potential is shifted from $r_{shift} = 0.0$ nm to r_{cut}, where as the LJ-potential is shifted from 0.9 nm, which means that starting from r_{shift} the potential functions are smoothly brought to zero at r_{cut} by adding a polynomial to the force function so that the derivatives are continuous at the boundary r_{cut}. The standard shift function in GROMACS is used for this purpose [25]. The shifted Coulomb potential mimics distance dependent screening.

A few special bead types are also introduced for special purposes, namely the anti-freeze and ring particles. Martini water has the propensity to freeze at 290 ± 5 K in a nucleation driven process quickly and irreversibly. This issue is solved by adding antifreeze particles, where the LJ-parameter σ is increased to 0.57 nm between antifreeze and normal water particles. Adding 10 % of anti-freeze particles into the system is enough to lower the freezing point below 250 K [24]. The ring particles extend the Martini model to small molecules containing a ring structure. The ring particles have reduced interaction size and strength to allow close packing of planes with out freezing.

Bonded interactions are described with weak harmonic potentials for the bond length and angle. For bonds the potential stands as

$$V_{bond}(r_{ij}) = \frac{1}{2} K_{bond}(r_{ij} - r^0_{bond})^2. \tag{3.6}$$

where $K_{bond} = 1250$ kJ mol^{-1}nm^{-2} is the force constant and $r_{bond} = 0.47$ nm is the equilibrium distance. The bond length may be varied if needed. Bonded beads are slightly closer to each other than non-bonded on average. The chain stiffness is

expressed in the cosine-type weak harmonic potential for the bond angles

$$V_{angle}(\theta_{ijk}) = \frac{1}{2} K_{angle}[\cos(\theta_{ijk}) - \cos(\theta^0_{ijk})]^2, \tag{3.7}$$

where θ_{ijk} is the angle between particles i, j and k and θ^0_{ijk} is the equilibrium angle at 180 ° by default. The small force constant $K_{angle} = 25$ kJ mol^{-1} rad^{-2} allows fluctuations up to 30° at the expense of one k_BT. LJ-interactions are excluded between nearest bonded neighbours, but not between second nearest. It has been shown [26] that with these parameters the configurational entropies and angle distributions of CG aliphatic chains compare well to atomistic models. In the case of cis-double bonds, the force constant needs to be changed to $K_{angle} = 45$ kJ mol^{-1} rad^{-2} and the equilibrium angle to 120°.

Martini force field can reproduce the partitioning free energies between polar water and several organic apolar environments accurately within 2 k_BT for a variety of different chemical compounds modeled with a single Martini bead. The good fit between experimental data and coarse grained simulation data applies to polar, non-polar as well as apolar molecules [24]. Particularly it is worth mentioning that the water/octanol partitioning is accurately generated since water/n-octanol partition ratio is used to predict the bioaccumulation of chemicals — toxins as well as drugs. The vaporization and hydration free energies on the other hand are systematically too high, but show the correct trend for both apolar and polar particle types. This is because the CG model underestimates the interaction energy of alkanes and overestimates the repulsion between water and alkanes [27]. Another difficulty for simulations with water is that the fluid state is not as stable as it should be with respect to the vapor or the solid phase at room temperature. The instability arises from the use of Lennard-Jones potential, which has a limited fluid range. This setback is not too problematic when simulating systems in the condensed phase, where the water/oil partitioning behavior is the most important.

The coarse-grained simulations are computationally efficient and can be as much as 1000 times faster to run than atomistic simulations. Two main factors contribute to the speed up. First, with simple soft potentials it is possible to use a large time step. Martini model is stable with time steps up to $dt = 40$ fs with no significant changes in the density or the free-energies of the system [6]. Second, due to lighter level of detail in the model the energy landscape is smoother and coarse grained systems sample the phase space much faster than atomistic systems. This is seen in the accelerated dynamics in the system. The particles are larger and friction caused by atomistic degrees of freedom has vanished. The speed up factor in the time scale for a variety of phenomena lies between two and ten, the average being four. As a first approximation the time axis can be scaled with a factor of four.

3.4 Polarizable Martini

The most crucial weakness of the Martini water model is the lack of polarizability. This leaves many essential interactions based on electrostatics out of the scope of simulations.

The polarizable Martini model aims to create a more realistic water model to use with Martini force field. Three criteria were used in the design: the dielectric constant of the model should be reasonably close to the one in bulk water while keeping the number density and partitioning free energies of the original Martini water. As the good reproduction of partitioning free energies for a wide range of solvents is at the heart of the Martini force field, the attempt to embed polarizability must not undermine the correct partitioning free energies.

The polarizable Martini uses four-to-one mapping, but now one water bead has three constituents. The model is illustrated in Fig 3.2. The neutral W particle in the center interacts with other particles only via the LJ-interaction. The charged particles WM and WP are tied to a distance $l = 0.14$ nm from the central particle and interact with other particles only through electrostatic interaction. The bond length l is chosen so that the charged particles stay inside the LJ-radius of the W particle. The LJ-distance parameter is kept at the same value as in original Martini ($\sigma = 0.47$ nm). The interaction between WM and WP of the same water is excluded and they are free to rotate around the central particle. WM and WP carry a charge of $q = \pm 0.46$. A harmonic potential is set between the charged particles with a spring constant $K_{angle} = 4.2$ kJ mol^{-1}rad^{-2} and equilibrium angle $\theta = 0$. This adjusts the distribution of the dipole moment. By setting the equilibrium angle at zero, one can be sure that the dipole moment vanishes in apolar environments. Each particle in the bead has a mass 24 amu totaling 72 amu which is the same as original Martini water.

The added electrostatic interaction increases the affinity of the beads as well as affects the partitioning free energies, and needs to be counterbalanced by reducing the LJ-interaction between the central W-particles. This self-interaction was reduced from $\epsilon_{LJ} = 5.0$ kJ mol^{-1} to $\epsilon_{LJ} = 4.0$ kJ mol^{-1} which corresponds to decreasing the interaction level from I to III in the Martini force field. Another consequence of the explicit treatment of polarizability is that the global dielectric constant $\epsilon_r = 15$ in the original Martini is reduced to $\epsilon_r = 2.5$ to create correct dielectric environment in hydrophobic areas. Adjustments need to be done also to the interaction levels between the W particle and other types of beads in the force field in order to correct the partitioning free energies. If the interaction levels between W and other beads are kept as they are in the standard Martini, ΔG_{part} will be too favorable. To fix this the interaction strength has been reduced by 5 %.

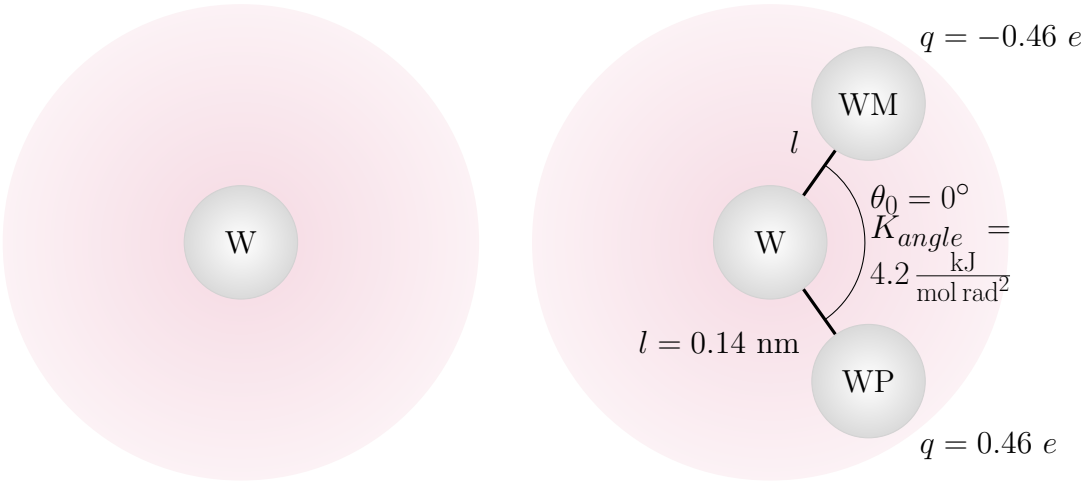


Figure 3.2: Illustrations of the coarse grained water models. Left: standard Martini bead. Right: polarizable Martini bead, where one particle has three constituents: the neutral W, WM (charge -q) and WP (charge +q)

.

The polarizable water model has a good semi-quantitative agreement with experiments and atomistic models. There are several compromises done in the model, but the gains in the improved electrostatics make them all worth it. Compared to the standard Martini, the density is increased by 4.7 % to 1043 kg m^{-3}, but the freezing point is lowered to 282 ± 2 K. The diffusion is slightly higher than in the standard or the experimental value $D_s = 2.5 \cdot 10^{-5}$ cm^2 s^{-1}. The dielectric constant is ∼4 % smaller than the experimental value of 75.6 debye at 300 K. The dielectric constant and density show the correct decreasing trend as function of temperature between 300–350 K as observed in experiments albeit the dependence on temperature is stronger in both cases. For the polar bead types the partitioning free energies are the same as in the standard Martini, but the values are smaller for apolar types and closer to experimental values. The distribution of the dipole moment is skewed and sharper than distribution obtained with atomistic SPC/E water model, but the semi-quantitative agreement is good. There is no certainty that the distribution of the atomistic model actually represents real water, so some reservations regarding the comparison have to be made. Never the less, the atomistic distribution is still the best approximation and the only to compare with. The maximum dipole moment of a single water bead is given by $2ql$ which yields 6.2 debye where as the maximum for four SPC/E water molecules is 9.4 debye. The polarizable Martini will inherently produce smaller dipole moments than atomistic models such as SPC/E.

The structure of water is the same for both models as judged by radial distribution functions in high ionic strength solution. Most significantly the structure of CG bilayers in conjunction with polarizable water is not changed, but the electrostatic potential across the membrane is improved a lot but still differs with results from

atomistic simulations. In atomistic simulations the surrounding water overcomes the negative charge of the lipid head groups and creates a negative overall potential inside the bilayer, where as the polarizable water produces too small a potential to generate a positive overall potential. The diffusion in lipid bilayers is slowed down compared to the standard Martini, but still remains in good agreement with experimental values. The electrostatics clearly slow down the movement. Phenomena like electroporation and ion translocation through a lipid bilayer can be successfully simulated with the polarizable water model.

From computational point of view the embedded electrostatics in polarizable Martini water brings out the question of long range interactions. It is observed that long range Coulomb interactions of polarizable water are well suited to be used with Particle Mesh Ewald technique (PME) as well as with shifted cut-off [28]. Therefore in this study, PME is used in all simulations. With PME the densities will be 0.6–1.1 % lower and diffusion slightly faster at elevated temperatures.

3.5 Effects of Coarse-graining on Dynamic Properties

The process of coarse-graining loses some information concerning the system, because the number of degrees of freedom is curtailed. This has effects in the thermodynamic properties as well as in the dynamic properties of the system. Let us examine theoretically the effects of coarse-graining in the case of four-to-one mapping taking ideal gas as a model system.

In general, the effects of coarse-graining can be considered to originate from two factors. First, the mapping itself reduces the number of particles in the system and hence effects the quantities, which depend directly on the number of particles, such as entropy or Gibbs free energy or even self-diffusion. When we average over particles it also produces smoother energy landscapes, which accelerates the dynamics of the system *regardless of the details of the mapping used* [6]. The second factor is the parametrization of the coarse-grained force field which determines the precise interactions between the particles. Naturally it is possible to gain different results by changing the force field parameters while keeping the mapping fixed. In order to compare the coarse-grained models with fine-grained ones both of these factors need to be considered in detail. There are two different speed up or mapping factors associated with these effects and in order to gain comparable results we need to know both of them. A simulation with a CG model happens in a simulation time/coarse-grained picture and a simulation with a fine-grained model happens in effective time/fine-grained picture, because for fine-grained model simulation time equals effective time. The factor associated with the mapping from coarse-grained to fine-grained shall be called the geometric coefficient and from simulations we gain another model-specific factor which is assigned to the mapping from simulation time to

effective time.

In the following a framework is developed where the contribution of the aforementioned factors to diffusion coefficients are characterised. The geometric factor can be derived theoretically while the force field factor is discovered from simulations. A 2 × 2 matrix presented in Table 3.4 clarifies the framework.

	Fine-grained		Coarse-grained
Simulation time	$A_{FG,sim}$	$\underset{\leftarrow}{\cdot c_m}$	$A_{CG,sim}$
	$\downarrow \cdot c_{ff}$		
Effective time	$A_{FG,eff}$		

Table 3.4: Framework for interpreting analysis results and making them comparable.

In the table A is the dynamic variable in different situations, c_m is the mapping coefficient and c_{ff} is the force field coefficient. The results from the CG simulations are in simulation time and when comparing the results to fine grained models, the conversion factor c_m is needed. The lower right corner is empty because coarse-grained effective time results are meaningless: there is no equivalent for them in the physical world. For all atom simulations $c_{ff} = 1$ and the simulation time matches the effective time. In the following the geometric coefficients for self-diffusion and collective diffusion are derived. Comparisons between the results from the simulations need to be done in the same picture that is in FG/simulation time. The geometric coefficients are needed to make this possible.

Self-diffusion

The trajectory of an ideal gas particle constitutes a random walk when it collides with other particles in the system. This is true also for the center of mass of four individual particles. If the individual particles move a distance $\Delta\mathbf{R}$ over a time t then the center of mass of four particles moves with

$$\mathbf{R}_{CG} = \frac{1}{4}\sum_{i=1}^{4}\Delta\mathbf{R}_i, \tag{3.8}$$

where $\Delta\mathbf{R}_i$ are the traveled distances of the individual particles. Considering the mean square displacement this converts to

$$R_{CG}^2 = \langle \mathbf{R}_{CG}\cdot\mathbf{R}_{CG}\rangle = \frac{1}{16}\left\langle \sum_{i=1}^{4}\mathbf{R}_i \cdot \sum_{j=1}^{4}\mathbf{R}_j \right\rangle \qquad = \frac{1}{16}\left(\sum_{i=1}^{4}\langle \mathbf{R}_i\cdot\mathbf{R}_i\rangle\right) \tag{3.9}$$

since the motions of ideal gas particles are uncorrelated, cross terms ($i \neq j$) in the dot product disappear and since the ideal gas particles are identical the sum can be

converted to multiplication by a scalar

$$\frac{4}{16}\langle \mathbf{R}_i \cdot \mathbf{R}_i \rangle = \frac{1}{4}R^2. \tag{3.10}$$

Inserting the above to the equation for self-diffusion (Eq. 2.8) it can be seen that the self-diffusion coefficient of the center of mass of four particles, which equals the coarse-grained particle, is one quarter of that for the individual particle.

$$D_s^{CG} = \frac{D_s}{4}. \tag{3.11}$$

Hence the mapping factor for self-diffusion is $c_{m,D_s} = 4$. Although the result is derived from ideal gas assumption and omits the effect of interactions, it is still a decent approximation. Studies by Bulavin *et al.* suggest that the collective effects to self-diffusion are significant [29]. Now we move on to derive the geometric coefficient in collective diffusion.

Collective Diffusion

To derive the relationship between coarse-grained and fine-grained collective diffusion coefficients, we can follow the same line of reasoning as in the case of self-diffusion. Let us observe the motion of the center of mass of four identical particles:

$$\sum_{i=1}^{N} \mathbf{R}_i(t) = \sum_{j=1}^{N/4} \sum_{k=1}^{4} \mathbf{R}_{jk}(t).$$

We rearrange the sum to make the centers of mass visible and make the same conclusions as before: the center of mass of four particles moves four times slower as the individual particles

$$\sum_{k=1}^{4} \mathbf{R}_{jk}(t) = 4\mathbf{R}_j^{CG}.$$

Inserting the above expression to the sum yields

$$\sum_{i=1}^{N} \mathbf{R}_i(t) = \sum_{j=1}^{N/4} 4\mathbf{R}_j^{CG}. \tag{3.12}$$

As mentioned before (Eq. 2.35), the Green-Kubo equation for a transport coefficient C_T can be written as

$$C_T = \frac{\beta}{V}\int_0^\infty \langle \dot{A}(t)\dot{A}\rangle dt, \tag{3.13}$$

where A is a microscopic dynamic variable, which is equivalent to writing

$$C_T = \frac{\beta}{V} \lim_{t\to\infty} \frac{1}{2t} \langle |A(t) - A(0)|^2 \rangle. \tag{3.14}$$

For collective diffusion D_c the dynamic variable is $A = \sum_N \mathbf{r}(t)$. Therefore, the equation for collective diffusion in Einstein form is

$$D_c = \frac{1}{dk_B T \rho N \kappa_T} \lim_{t\to\infty} \frac{1}{2t} \left\langle \left| \sum_N \mathbf{r}(t) - \sum_N \mathbf{r}(0) \right|^2 \right\rangle. \tag{3.15}$$

The sum over fine-grained particles can be replaced with the coarse-grained sum:

$$\begin{aligned} D_c &= \frac{1}{dk_B T \rho N \kappa_T} \lim_{t\to\infty} \frac{1}{2t} \left\langle \left| \sum_N \mathbf{r}(t) - \sum_N \mathbf{r}(0) \right|^2 \right\rangle && (3.16) \\ &= \frac{1}{dk_B T \rho N \kappa_T} \lim_{t\to\infty} \frac{1}{2t} \left\langle \left| \sum_{j=1}^{N/4} 4\mathbf{R}_j^{CG}(t) - \sum_{j=1}^{N/4} 4\mathbf{R}_j^{CG}(0) \right|^2 \right\rangle && (3.17) \\ &= \frac{16}{dk_B T \rho N \kappa_T} \lim_{t\to\infty} \frac{1}{2t} \left\langle \left| \sum_{j=1}^{N/4} \mathbf{R}_j^{CG}(t) - \sum_{j=1}^{N/4} \mathbf{R}_j^{CG}(0) \right|^2 \right\rangle && (3.18) \\ & && (3.19) \end{aligned}$$

The last line is just the definition of collective diffusion for the coarse-grained particles (or centers of mass). Thus it can be written

$$D_c = 16 D_c^{CG}. \tag{3.20}$$

We see that the speed up coefficient from the four-to-one mapping is $c_{m,D_c} = 16$. Again the same limitations apply as with self-diffusion.

In this chapter the discussion has been on the practical elements of modeling water and getting to know the models used in this thesis. All of the models used are derived using empirical approach, which aims to produce models suitable to a wide range of applications. The atomistic model SPC has strengths in simplicity and providing an overall good match with actual water. The coarse-grained models are semiquantitative in nature meaning that the properties and behaviour of the model match real water when painting with a broad brush. There is an important and deep relation between the FG and CG models when it comes to dynamic properties and to get a concrete grasp of the connection the speed up coefficients of self-diffusion and collective diffusion were derived. This connection is central in the underpinning of interpreting the results and final conclusions. Viscosity is omitted here, because the case for viscosity is much more complicated and deriving the speed up coefficient between CG and FG models is beyond the scope of this thesis. After gaining an

overview of the simulation and analysis methods we are well equipped to look at the actual results and understand in the light of the theory presented in this and the previous chapter.

4. METHODS

It is relatively simple to probe a sample with some input and measure the response, but to discover the molecular mechanism behind the response is difficult — if not right out impossible. The measured signal is something macroscopic, while molecules and their interactions are microscopic. With computer simulations it is possible to fill this gap and explore the motions of single molecules and their interactions on the atomic scale. Simulations try to enlighten the question how the molecules interact by approximating the molecular interactions with an empirical force field. Only during the last two decades the experimental techniques, such as single-molecule atomic force microscopy, have managed to study the properties of single molecules. Next, the basic concepts of the simulation technique used in this thesis, Molecular Dynamics, are explained as well as the analysis methods to computationally determine the dynamic properties discussed in Chapter 2.

4.1 Molecular Dynamics

The two most widely used simulation techniques for studying atomic ensembles are Monte Carlo (MC) and Molecular Dynamics (MD). MC methods use random sampling to compute the desired quantities of a system and MD uses the classical equations of motions to determine the time evolution of the system at hand. The biggest difference between the techniques is that MC cannot produce the time evolution of the simulated system, whereas this is in the core of the MD method. Monte Carlo methods can be applied in finance, physics, mathematics, telecommunications and biology, but Molecular Dynamics is restricted to physical systems. In this thesis the focus is on the dynamics and time scales, so our method of choice is MD. Before jumping into the world of MD, a basic understanding of some concepts from statistical mechanics is useful.

Theoretical Considerations and Basic Concepts

Imagine the most simple statistical system: a particle in a box. If we would like to completely and uniquely capture the state of such a (classical) single-particle system, we would need to know the location and the momentum of the particle. These can be described with three spatial coordinates and three momentum coordinates, which create a six dimensional space, in which each point describes one state of the

system. As the particle moves inside the box, so does its *phase point* move in the six dimensional space. Similarly, to describe the state of a system consisting of N particles we need six dimensions for each particle and in total a $6N$-dimensional space, which is called *the phase space* (or Γ-space). The $3N$ momentum coordinates and $3N$ spatial coordinates are collectively denoted by ($\mathbf{p}$,$\mathbf{q}$), respectively. Now we see the benefit of this notation: we can simplify the statistical system into one point and follow its movement instead of following all the N particles separately. The phase space is a central concept in statistical mechanics and also in computer simulations.

The N-particle system moves through the phase space as the system evolves over time. The dynamics of this motion are completely contained in the Hamiltonian $H(\mathbf{p},\mathbf{q})$ and the Hamiltonian equations of motion

$$\frac{d\mathbf{q}_i}{dt} = \frac{\partial H(\mathbf{p},\mathbf{q})}{\partial \mathbf{p}_i} \quad \text{and} \quad \frac{d\mathbf{p}_i}{dt} = -\frac{\partial H(\mathbf{p},\mathbf{q})}{\partial \mathbf{q}_i} \quad \text{where} \quad i = 1, \cdots, N. \tag{4.1}$$

The phase point representing the particles is not free to wander around as it wishes, because the different statistical ensembles impose different restrictions to the trajectory. How then do different statistical ensembles differ from one another in phase space? The microcanonical ensemble has a fixed number of particles, volume and energy (NVE-ensemble) which means $H(\mathbf{p},\mathbf{q}) = E$. In phase space this means that NVE-ensembles constitute a constant energy surface, where the phase point of the system moves, and only the points on the surface are accessible to the system. In the canonical ensemble energy is allowed to fluctuate, but temperature is kept fixed (NVT-ensemble). Now all the points in phase space are attainable to the system and the probability of visiting a phase point is proportional to the Boltzmann factor

$$e^{-A/k_BT}, \tag{4.2}$$

where A is the Helmholtz free energy, k_B is the Boltzmann constant and T is the absolute temperature. States with high energy have a small probability to be visited due to the exponential nature of the Boltzmann factor. The most common ensemble in MD simulations is the NPT -ensemble, also known as the isothermal-isobaric ensemble, where all phase points are accessible, but the probability is proportional to e^{-PV/k_BT}, where PV is the product of pressure and volume called *the grand potential*. In the thermodynamic limit all ensembles produce the same results [9].

A simulation system mimics the real system it represents by sampling the same regions of the phase space as the real system, but a simulation has only a finite number of steps to do this. Thus one measure for the efficiency of the model is the extent it can sample the phase space in a limited time. In a grand canonical ensemble

the probabilities of visiting a state are proportional to the Boltzmann factor, which means that states with low probability (ie. high energy) are not sampled properly [30]. This is to be considered, when interpreting the results from a simulation.

The Molecular Dynamics Algorithm

Computer simulations have many similarities to conducting experiments: first, we need to choose the materials or models and prepare the sample(s), then do the measurement or run the simulation, and finally analyze and interpret the results. A schematic flow chart of the MD simulation algorithm is depicted in Fig 4.1 and a more realistic flow chart is depicted in Fig 4.1, which is the actually used flow in the simulation package GROMACS 4.5.1. The basic algorithm (4.1) is simple as it should be for easy implementation. Real flow shows more details and highlights some of the computational aspects of conducting an efficient simulation. For instance, the long range electrostatic interactions are calculated efficiently on a separate PME node.

Preparing the sample in a computer simulation means defining the topology of the system. Topology includes the initial positions and velocities of the particles, which constitute the molecules, and the interactions between the particles. The interactions are defined in the force field of the model as discussed in the Models Chapter. The initial conditions must be reasonably close to the equilibrium conditions (or in the high probability region of phase space) so that we can be sure that the system will find equilibrium during the restricted simulation time. The correct distribution of velocities for the particles is generated from the Maxwell-Boltzmann distribution:

$$p(v_i) = \sqrt{\frac{m_i}{2\pi k_B T}} e^{-\frac{m_i v_i^2}{2k_B T}}, \tag{4.3}$$

where m_i and v_i are the mass and velocity of the particle, respectively, k_B is the Boltzmann constant and T is the absolute temperature.

The simulation system is assembled by randomly locating the molecules in the simulation box, which may place certain atoms too close to each other. Before the actual simulation the energy of the system is minimized by moving the particles in minute displacements and trying to find the local energy minimum in the vicinity of the initial configuration. Energy minimization is not enough, and a short equilibration simulation is run before the actual production simulation.

The core of the simulation is solving the equations of motion with short time intervals (or time steps) and moving the particles to new locations based on the

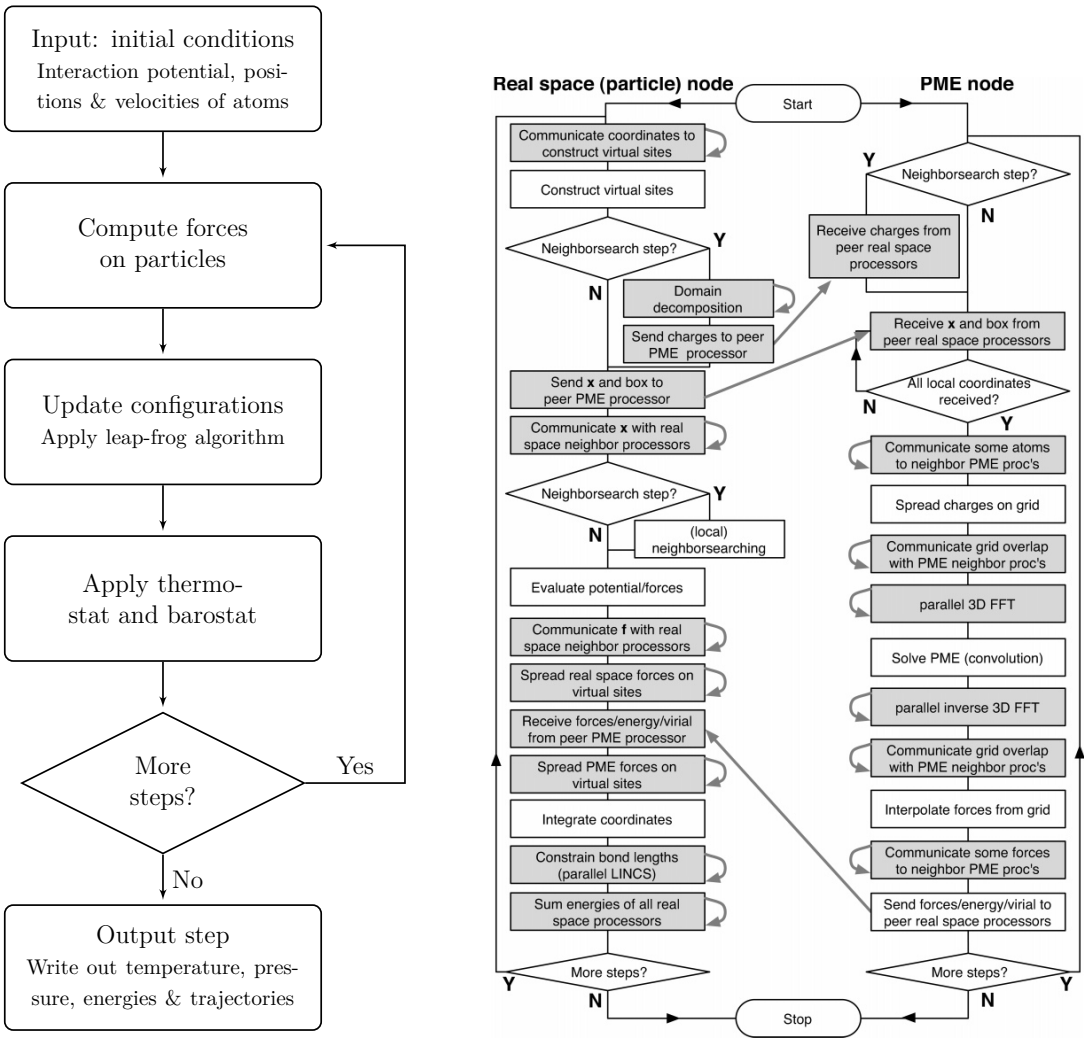


Figure 4.1: On the left, a schematic flowchart of MD algorithm and on the right, actual algorithm used in GROMACS, from [25]

forces acting on them:

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, \quad i = 1, \ldots, N, \tag{4.4}$$

where $\mathbf{F}_i$ is the resultant force on particle i, m_i and $\mathbf{r}_i$ are the mass and location of particle i. The resultant force is obtained from the potential

$$\mathbf{F}_i = -\frac{\partial U}{\partial \mathbf{r}_i}. \tag{4.5}$$

In Gromacs the time integration is implemented with *the leap frog* –algorithm, where the locations and the velocities are solved in turns every half step:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i\left(t + \frac{\Delta t}{2}\right)\Delta t \tag{4.6}$$

$$\mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}_i\left(t - \frac{\Delta t}{2}\right) + \frac{\mathbf{F}_i(t)}{2m_i}\Delta t, \tag{4.7}$$

where $\mathbf{v}_i$ and $\mathbf{r}_i$ are the velocity and location of particle i, Δt is the time step, t the current time, m_i and $\mathbf{F}_i(t)$ are the mass and the force acting on particle i. The leap frog algorithm is a version of the Verlet algorithm and it possesses many good features for an algorithm to solve the Newton's equations of motion. It is time reversible, low on memory consumption, and the error is of third order. The velocities and locations of all the particles in the system are recorded at regular intervals and this produces the output of the simulation.

There are several technical details that still need to be taken into account when running a simulation. The simulation box has a finite size, which raises the question what happens at the boundary of the simulation box. Handling the long range interactions demands combining physical accuracy and computational feasibility. In the NPT-ensemble, controlling the pressure and temperature are important tasks and require specific techniques. These technical issues are addressed in the following.

Periodic Boundary Conditions

Due to the finite size of the simulation box some boundary conditions need to be imposed. The typical solution is the so called *periodic boundary condition* where a particle leaving the box from left immediately re-enters the box from right. Likewise passing through the roof of the simulation box results in entering the box back from the ceiling. This corresponds to replicating the simulation box infinitely in every direction. The concept of periodic boundary conditions is illustrated in Fig 4.2. When calculating the forces, the program needs to know the closest neighbours to a given particle and the periodic boundary conditions have some ramifications when determinig this when a particle is close to the edge of the simulation box. Namely,

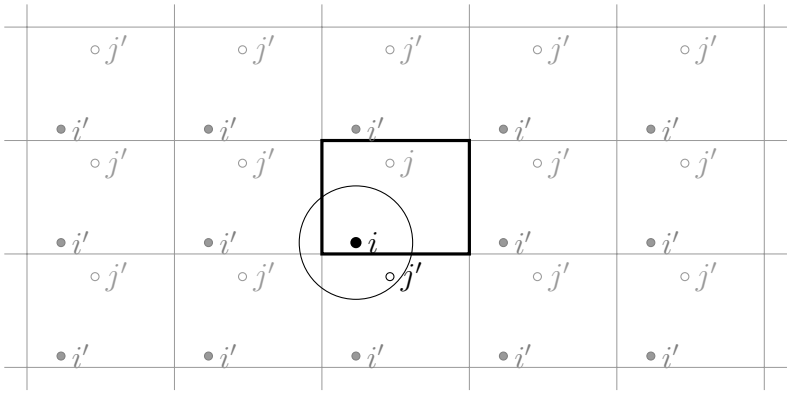


Figure 4.2: Periodic boundary conditions consider only the nearest neighbour whether it is located in the actual simulation box or its periodic image

the closest neighbours can be also found from the adjacent copy of the simulation box and not just in the one where the particle in question is located.

Long Range Interactions

Vast majority of the computational effort is consumed in the calculation of forces exerted on the particles. Simulations typically employ a cut-off scheme to decrease the computational cost of computing the forces on particles by omitting the interactions between particles beyond a certain distance. This is very problematic for the long range electrostatic forces, because it causes the long tail contribution to be neglected, which increases the error in the simulation. In the worst case the neglect leads to unphysical artifacts at the cut-off distance, such as overestimating the structure of the liquid or underestimating the repulsion between charged groups. There is a clear need for advanced methods to include the long range interactions with reasonable computation, and this call has been answered with a multitude of methods, such as Ewald summation [31], reaction field [32], Particle-Mesh Ewald (PME) [33, 34] and Particle-Particle Mesh Ewald [35, 36]. In the following, the focus is on PME method, because it is the method used in the simulations in this thesis.

The PME method is accurate and computationally efficient. The electrostatics are important in the polarizable Martini model and therefore they need to be computed properly in the simulation. The fundamental problem of computing the electrostatic energy is the slow convergence of the total electrostatic energy sum when it includes the periodic images

$$U_e = \frac{f}{2} \sum_{n_x} \sum_{n_y} \sum_{n_z *} \sum_{i}^{N} \sum_{j}^{N} \frac{q_i q_j}{\mathbf{r}_{ij,\mathbf{n}}}, \tag{4.8}$$

where $(n_x,n_y,n_z) = \mathbf{n}$ is the box index vector and the star indicates that at $i = j$

the terms should be omitted when $\mathbf{n} = \mathbf{0}$. The sum considers the distance $\mathbf{r}_{ij,\mathbf{n}}$ as the real distance between charges and not just the minimum image distance.

The idea in PME is to convert the sum V into a constant and two quickly converging sums: a short range part computed in the real space and a long range part computed in reciprocal space using Fourier transform. The short range part up to r_{cut} is computed with a modified Coulomb's law [25]

$$U_{mod}(r) = U_c(r)\mathrm{erfc}(\zeta r), \tag{4.9}$$

where $U_c(r)$ is the Coulomb potential and

$$\mathrm{erfc}(\zeta r) = \frac{2}{\pi^{1/2}} \int_{\zeta r}^{\infty} \exp(-t^2)dt \tag{4.10}$$

is the complementary error function and ζ is a parameter to determine the relative weight between the real space sum and the reciprocal space sum. In reciprocal space the charges are interpolated onto a discrete mesh or a grid, on which the Fast Fourier transform (FFT) can be efficiently applied to obtain the long-range energy term. The inverse Fourier transformation is then applied to calculate the potential at the grid points and with the interpolation factors the forces on each particle are recovered.

PME is a further development of the Ewald summation method. The advantage of PME is the efficient application of FFT making the PME algorithm of order $N\log(N)$ in contrast to the Ewald sum which is of order N^2 [33]. The second advantage, which is particularly important in this thesis, is that it does not cause artifacts in pressure fluctuations [37].

Pressure and Temperature Coupling

Biological phenomena and often experiments happen in conditions of constant temperature and pressure. Since isobaric-isothermal conditions match the physiological conditions, biophysical simulations are most often run with the NPT-ensemble, which requires a thermostat and a barostat in the simulation to keep the temperature and pressure constant. In the simulations in this thesis, the velocity rescaling thermostat [38] and Berendsen barostat [39] are used.

The temperature T in a system is defined through its kinetic energy E_{kin} using

the equipartition theorem [25]

$$E_{kin}(t) = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2(t) = \frac{1}{2} N_{df} k_B T \tag{4.11}$$

$$T(t) = \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{k_B N_{df}} \tag{4.12}$$

where m_i and $v_i(t)$ are the mass and velocity of the particle i at a time t and N_{df} is the number of degrees of freedom in the system. The velocity rescaling thermostat controls the temperature by rescaling all the velocities of the particles with a random factor drawn from the canonical distribution. The equation for the dynamics of E_{kin} is

$$dE_{kin} = (E_{kin}^0 - E_{kin})\frac{dt}{\tau} + \sqrt{\frac{E_{kin}E_{kin}^0}{N_{df}}}\frac{dW}{\sqrt{\tau}} \tag{4.13}$$

where E_{kin}^0 is the target kinetic energy, E_{kin} is the observed kinetic energy, τ is the time constant, N_{df} is the number of degrees of freedom and dW is Wiener noise. Without the latter stochastic term this is equivalent to the Berendsen thermostat. In the velocity rescaling thermostat temperature oscillations decay exponentially far from equilibrium. The thermostat has one important advantage over the Berendsen thermostat: the correct canonical ensemble is produced.

The other constant in a NPT-simulation, the pressure, is computed from the trace of the pressure tensor $\mathbf{P}$

$$\mathbf{P} = \frac{2}{V}(E_{kin} - \mathbf{\Xi}), \tag{4.14}$$

where V is the volume of the system, E_{kin} is the kinetic energy and Ξ is the virial tensor

$$\mathbf{\Xi} = -\frac{1}{2}\sum_{i<j} \mathbf{r}_{ij} \otimes \mathbf{F}_{ij}. \tag{4.15}$$

The scalar pressure is computed as the trace of the pressure tensor

$$P = \frac{1}{3}\sum_{i=j} \mathbf{P}_{ij}. \tag{4.16}$$

The Berendsen barostat controls the pressure by a weak coupling to an external 'pressure bath.' The rate of change of the pressure is proportional to the fluctuation

from the target pressure and the characteristic time constant:

$$\frac{\mathbf{P}}{t} = \frac{\mathbf{P}^0 - \mathbf{P}}{\tau_p}, \tag{4.17}$$

where $\mathbf{P}^0$ is the target pressure, $\mathbf{P}$ is the current pressure and τ_p is the time constant chosen by the user. The simulation box vectors and the atomic coordinates are scaled every n_{PC} steps with a matrix μ given by

$$\mu_{ij} = \delta_{ij} - \frac{n_{PC}\Delta t}{3\tau_p}\kappa_{ij}(P^0_{ij} - P_{ij}(t)). \tag{4.18}$$

Here δ_{ij} is the Kronecker delta, κ_{ij} is the isothermal compressibility, Δt is the time step, τ_p is the pressure coupling time constant and P^0_{ij} is the target pressure.

4.2 Analysis Methods

For each model three dynamic properties are analysed from the simulations. There is a variety of possible methods to analyse the properties and thus care needs to be taken in choosing the most suitable one. Only for one of the properties, self-diffusion, there exists a ready-made tool in GROMACS simulation package. For the other two, viscosity and collective diffusion, the analysis codes are self-made. In a comparative manner short descriptions of different methods are included and the rationale for the chosen methods is presented as well as the technical details of the methods.

Self-diffusion

Self-diffusion, or tracer diffusion, is calculated in a standard way using the Einstein equation [12] and averaging over all the particles in the system:

$$D_s = \lim_{t\to\infty} \frac{1}{6t}\langle[\mathbf{r}(t - t_0) - \mathbf{r}(t_0)]^2\rangle, \tag{4.19}$$

where $\mathbf{r}$ are the coordinates of the centers of mass of the particles at times $t - t_0$ and t_0. Another method would be to use the Green-Kubo formalism as introduced in Chapter 2, but it converges slowly. The GROMACS tool `g_msd` implements the Einstein equation.

Collective Diffusion

Collective diffusion, which describes the collective density fluctuations in the system, can be computed using the Green-Kubo formula (Eq. 2.36), but since the correlation time is very short it is difficult to get an accurate estimate for the collective diffusion

factor. Instead, the *density-fluctuation method* [40] is used for the computation as explained in Chapter 2.

The computational effort required to compute collective diffusion with the same accuracy as self-diffusion is much larger, because D_c is a collective property of the system and all the particles in the system are required to obtain one sample for the evaluation of D_c. This is at variance to self-diffusion, where we get N samples from the simulation.

In order to calculate the local density the simulation box is divided into cubical subunits, which have dimensions approximately matching the particle size as defined through the van der Waals -radius of the coarse grained particles. Local density is defined as the number of particles inside the subunit divided by the volume of the subunit. The size of the subunit is an important choice. The local (number) density would be ill-defined if the subunits were smaller than the particle size. On the other end, the subunits cannot be too large because it makes it more difficult to measure the density fluctuations.

The density fluctuation is discovered by subtracting the average density from the local density, after which the Fourier transform is applied. From the several wavelengths given by the FFT-algorithm, the two longest are chosen in order to stay in the hydrodynamic regime of long times and small k. For simplicity the local density is calculated in each dimension separately. A single autocorrelation function is computed from the three dimensional density fluctuations, and an exponential function of the form

$$y = a \exp^{bx} \tag{4.20}$$

is fitted to the data. The autocorrelation function is calculated using `g_analyze` from the GROMACS simulation package [25]. The collective diffusion coefficient is extracted from the exponential. The one-dimensional autocorrelations would give the diagonal elements of the collective diffusion tensor $\mathbf{D}_c$ which can be combined to give the isotropic D_c

$$D_c = \frac{1}{3} Tr\ \mathbf{D}_c. \tag{4.21}$$

Now `g_analyze` combines the three dimensions automatically and only a single fit is needed to get the collective diffusion coefficient D_c.

Shear Viscosity

The shear viscosity can be easily obtained from experiments and used to compare the molecular models to reality. Since majority of the properties used in comparison are static or thermodynamic, shear viscosity and self-diffusion are valuable in shedding light into the dynamics. There are several ways to calculate the liquid's viscosity from MD simulation. One can use equilibrium methods, such as pressure

or momentum fluctuations [41], or non-equilibrium methods, such as periodic shear flow [42] or sliding boundary condition method with SLLOD algorithm [43]. A good survey of the performance of the different methods is given by Hess [37].

The pressure fluctuation method uses the intrinsic pressure fluctuations of the off-diagonal elements of the pressure tensor in the Green-Kubo equation to calculate the viscosity. It relies most closely on the definition of shear viscosity, but it converges slowly, because the pressure fluctuates strongly during simulation. It is also sensitive to the long-range interactions and to the method used for handling the long-range interactions, which have significant influence on the pressure fluctuations [37].

The momentum fluctuation method is based on the transverse-current autocorrelation functions (TCAFs) and is wave-length dependent [41]. In the momentum fluctuation method the decays of the transverse-currents are determined at different wave-lengths. To obtain the shear viscosity at certain wave-length, a fit to the analytic TCAF is needed. Then, the viscosities at different wave-lengths are extrapolated to zero-wave length to get the proper shear viscosity. There is some uncertainty in the extrapolation and also in the analytic fit.

The afore mentioned methods utilise the internal fluctuations of the system at equilibrium, whereas another approach is to impose an external perturbation on the system and calculate the viscosity from the response of the system in a non-equilibrium simulation. One must be careful, however, not to take the system too far from equilibrium, since this will effect the viscosity calculations. Here two common non-equilibrium methods are considered: the SLLOD algorithm and the periodic perturbations method. The SLLOD algorithm creates a Couette flow, that is a laminar flow between two plates, which is achieved by modifying the equations of motion of the particles. This results in a linear velocity profile and the shear viscosity is obtained from the off-diagonal component of the pressure tensor. The result of this procedure is that the equations of motion in the SLLOD algorithm are non-Hamiltonian: the induced velocities generate forces –not the other way around. This is unnatural and the preferred method in this work is *the periodic perturbations method* (PP-method).

The periodic perturbation method offers high signal-to-noise ratio and is computationally efficient [37]. In short the method relies on applying an external periodic force in the x-direction to all the particles in the system and calculating the viscosity from the response to this shearing force. Next, the equation for viscosity is derived in the PP-method starting from the basic equation of hydrodynamics. The Navier-Stokes equation stands as

$$\rho\frac{\partial \mathbf{u}}{\partial t} + \rho(\mathbf{u}\cdot\nabla)\mathbf{u} = \frac{\rho}{m}\mathbf{F}(\mathbf{r},\mathbf{t}) - \nabla p + \eta\Delta\mathbf{u} \tag{4.22}$$

where ρ is the density of the system, $\mathbf{u}$ is the velocity field, p is the pressure and η is the viscosity. The external force $\mathbf{F}(\mathbf{r},\mathbf{t})$ is chosen such that it is zero in the y and z directions and the value in x direction depends on z. Furthermore, to achieve a steady-state in the simulation, the force will also be time-independent. The equation 4.2, thus, simplifies to a one dimensional case:

$$\frac{\rho}{m}F_x(z) + \eta\frac{\partial^2 u_x(z)}{\partial z^2} = 0. \tag{4.23}$$

The external force profile needs to be smooth, so that it generates a smoothly varying velocity profile and in order to keep the local shear rates small. The force has to be large enough to create a steady flow in to the system, but not so large as to take the system far from equilibrium. For a smoothly varying function the cosine is chosen:

$$F_{i,x}(z) = m_i\mathcal{A}\cos(kz), \tag{4.24}$$

where $\mathcal{A}$ is the amplitude of the force and k is the wave-index number:

$$k = \frac{2\pi}{l_z}, \tag{4.25}$$

where l_z is the height of the simulation box. Since Navier-Stokes equation is not valid for microscopic length scales, the wavelength has to be at least an order of magnitude larger than the typical length scale of the system. This is guaranteed by making the height of the box an order of magnitude larger than the size of a molecule in the simulation [37]. The simplified Navier-Stokes is readily solved with the chosen external force. With initial condition $u_z(x) = 0$ at $t = 0$, the generated steady-state velocity profile is [37]

$$u_x(z) = \mathcal{V}(1 - e^{-t/\tau_r})\cos(kz) \tag{4.26}$$

$$\mathcal{V} = \mathcal{A}\frac{\rho}{\eta k^2} \tag{4.27}$$

where τ_r is the macroscopic relaxation time of the liquid given by $\tau_r = \frac{\rho}{\eta k^2}$. By measuring $\mathcal{V}$ from the simulation, the viscosity can be calculated using Eq. 4.26. In a simulation the instantaneous $\mathcal{V}(t)$ is defined as

$$\mathcal{V}(t) = 2\frac{\sum_{i=1}^{N} m_i v_{i,x}(t)\cos(k r_{i,z}(t))}{\sum_{i=1}^{N} m_i}, \tag{4.28}$$

where $v_{i,x}$ is the velocity in x-direction, $r_{i,z}$ is the z-coordinate and m_i is the mass of atom i. The estimated time for the full development of the velocity profile is $t = 5\tau_r$, after which it is possible to start measuring the average $\mathcal{V}$.

The larger one chooses the amplitude of the external force $\mathcal{A}$ to be the larger the amplitude of the velocity profile $\mathcal{V}$ is, and the better the statistics for determining viscosity. However, a large amplitude inflicts high shear rates and the system is not able to relax resulting in too low values for the measured viscosity. The upper limit for $\mathcal{A}$ is determined by the maximum shear rate:

$$s_{max} = max_z \left| \frac{\partial v_x(z)}{\partial z} \right| = \mathcal{A}\frac{\rho}{\eta k}. \tag{4.29}$$

The shear rate has units of inverse time. When inverse of s_{max} is larger than the typical relaxation time, the system is able to relax. For most liquids the correlation time used is the rotation correlation time and it has values around 10 ps [37]. The typical values for $\mathcal{A}$ are in the range 0.1 – 0.0025 nm ps^{-2} [28, 37].

The external force field is doing work on the system, which heats up the system. The generated heat is removed by a coupling to the thermostat, but since the coupling is not immediate the real temperature will be slightly lower than the observed temperature of the system. For the correct calculation of viscosity, the velocity profile should not, however, be coupled to the the heat bath and should be excluded from the kinetic energy [37]. For the Berendsen thermostat the shift in temperature is

$$T_s = \frac{\eta \tau_T}{2\rho C_v} s_{max}^2, \tag{4.30}$$

where τ_T is the Berendsen thermostat coupling time, C_v is the heat capacity, ρ is the density and s_{max} is the maximum shear rate. For water the temperature shift is typically ~ 0.2 K.

The periodic perturbation method gives an error in the viscosity

$$\sigma_\eta = \frac{2}{s_{max}} \sqrt{\frac{k_B T \eta}{t_a V}}, \tag{4.31}$$

where t_a is the time over which $\mathcal{V}(t)$ is averaged and V is the volume of the system. Error decreases with square root of time and volume (ie. number of particles). On the other hand, CPU-time increases linearly with the number of particles and thus the CPU-time needed for a certain level of accuracy is independent of system size [37].

4.3 Simulations in This Thesis

More than 50 simulations in total were run to analyse the dynamic properties of the three water models. Two different simulation boxes were used for diffusion and

Table 4.1: Simulation systems and simulation box sizes at the end of equilibration simulation.

System	# of particles	Box size in diffusion simulations (nm)	Box size in viscosity simulations (nm)
SPC	8192	6.33 × 6.33 × 6.33	4.07 × 4.07 × 15.41
Martini	2048	6.27 × 6.27 × 6.27	3.91 × 3.91 × 15.54
Polarizable Martini	2048	6.21 × 6.21 × 6.21	3.93 × 3.93 × 15.51

viscosity calculations. The simulation systems are described in Table 4.1, where the box sizes are only indicative, since the box size varies during the simulation and between different simulations in accordance to temperature.

The required number of molecules were randomly positioned in the simulation box using standard tools in GROMACS followed by energy minimization and an equilibration simulation. For SPC water, the equilibration simulation was 1 ns and for Martini water and polarizable Martini the equilibration runs were 2 ns all of them using Berendsen barostat and V-rescale thermostat. The reference temperature was 310 K with a time constant $\tau_T = 4.0$ ps, and reference pressure was 1.0 bar with a time constant $\tau_p = 2.0$ ps. For the Coulomb interactions PME treatment was used with a cut off radius $r_{cut} = 1.2$ nm. Standard time steps were used for all models.

In the analysis simulations the electrostatic treatment, barostat and thermostat along with the associated time constants were kept unchanged in all simulations whereas the durations varied from 5 ns to 10 ns with atomistic simulations, and from 10 ns to 100 ns with coarse grained simulations. Analysis simulations for collective and self-diffusion were the longest: 10 ns for SPC water and 100 ns for the coarse grained models. Long simulations ensure good statistics and provides reliability to the results. The temperature dependence of the dynamic variables were examined with simulations in the temperature range 300 K to 350 K with 10 K intervals. Simulated annealing was used to reach the target temperature in 200 ps. The systems were allowed to equilibrate properly in the analysis simulations by omitting the first nanosecond in SPC simulations and the first 10 to 25 nanoseconds in CG simulations.

The analysis simulations for viscosity were significantly shorter and ranged from 3 ns to 10 ns. It has been shown that a 2 ns simulation can produce good statistics if the system size is large enough [28, 37]. The analysis simulation for the SPC model was 5 ns and 10 ns for the coarse grained models. Six different acceleration amplitudes were used in coarse grained simulations: 0.01, 0.005, 0.004, 0.0025, 0.002 and 0.001 nm ps^{-2} to study the effect of the acceleration amplitude.

5. RESULTS AND DISCUSSION

In this chapter, the dynamic properties are used to characterize the coarse grained models, and a comparison to atomistic models is made. In particular the temperature dependence of the dynamic quantities is analysed in detail. Starting with self-diffusion and collective diffusion we move on to viscosities. The attempt is to understand the time scales in different dynamic properties.

5.1 Self-Diffusion

Self-diffusion coefficient is the standard measure of the dynamics of a system and thus the discussion is started with the self-diffusion coefficient at different temperatures as presented in Fig. 5.1. Here the results for CG models are in effective time and for individual water molecules making them comparable to SPC and experimental results. The numerical values are presented in Table 5.1. The figure shows that the atomistic SPC water has a much stronger response to temperature than coarse grained counterparts, and the self-diffusion coefficient more than doubles from $4.4 \cdot 10^{-5}$ cm^2 s^{-1} to $10.0 \cdot 10^{-5}$ cm^2 s^{-1} when the temperature is raised from 300 to 350 K. The coarse grained models behave almost identically with a strictly linear response polarizable Martini having larger values by $3.71 \cdot 10^{-5}$ to $5.73 \cdot 10^{-5}$ cm^2 s^{-1}. The speed up is clearly visible in the results as well as the fact that it does not stay constant over the whole temperature interval.

Table 5.1: Self-diffusion coefficients between temperatures 300–350 K. The experimental results have been intrapolated to match the simulation temperatures. Experimental data from [44]

	$D_s[10^{-5}$ cm^2 s^{-1}]			
T [K]	SPC	P-Martini	Martini	Exp
300	4.40 ± 0.01	10.11 ± 0.12	6.40 ± 0.24	2.60
310	5.16 ± 0.02	11.89 ± 0.04	7.63 ± 0.12	3.22
320	5.91 ± 0.12	13.26 ± 0.40	9.01 ± 0.04	3.98
330	7.24 ± 0.15	15.52 ± 0.32	10.26 ± 0.16	4.77
340	8.23 ± 0.31	17.80 ± 0.68	11.64 ± 0.40	5.65
350	10.02 ± 0.80	18.80 ± 0.60	13.07 ± 0.16	6.58

The ratios of experimental D_s to coarse grained D_s have been calculated in Table 5.2 where we can see that the speed up diminishes with increasing temperature.

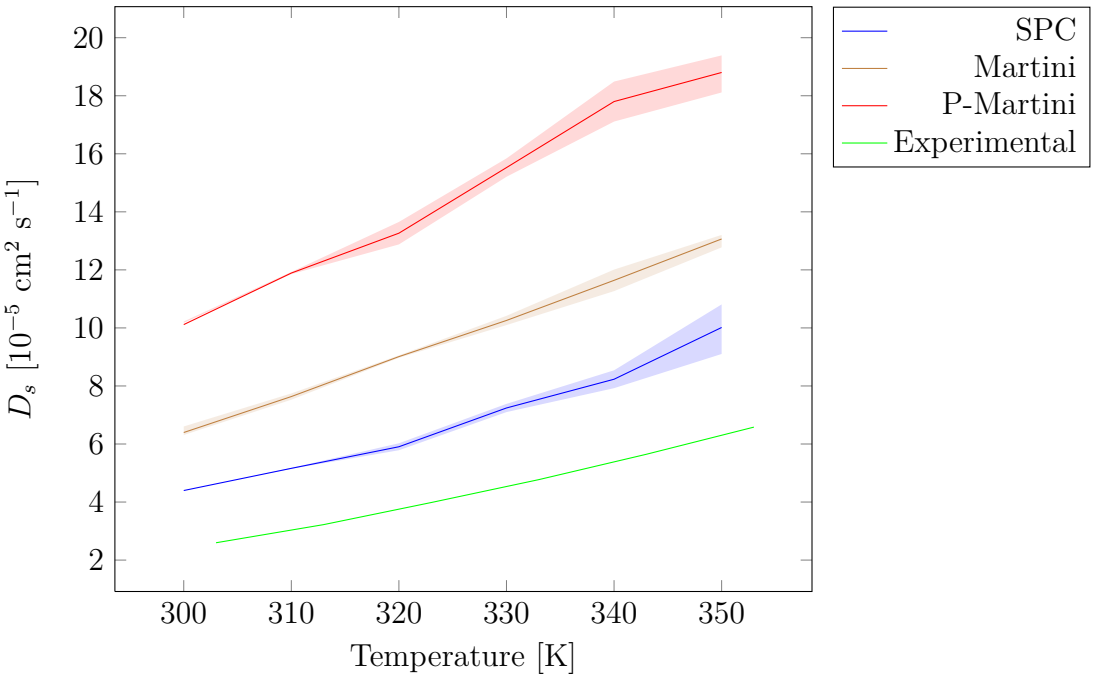


Figure 5.1: The self-diffusion of different models in the temperature range 300–350 K including error limits. Experimental data from [44]

Table 5.2: The speed up factors for self-diffusion between temperatures 300–350 K. Experimental data from [44]

T [K]	Speed up factors		
	SPC/Exp	P-Martini/Exp	Martini/Exp
300	1.792	2.601	4.112
310	1.712	2.533	3.945
320	1.575	2.404	3.538
330	1.598	2.264	3.427
340	1.533	2.167	3.314
350	1.579	2.059	2.964

For the Martini model the change in speed up is 28 % over the whole temperature interval, which is significant. For polarizable Martini the change is 21 %.

5.2 Collective Diffusion

There are no previous thorough studies on collective diffusion of water through MD simulations. In order to examine the reliability and accuracy of the density-fluctuations method for computing the collective diffusion (as described in Chapter 3), it was tested against a few essential parameters: the effect of simulation length and time interval between consecutive frames in the simulation, and the impact of sub-unit length. The simulation length an important variable in order to see how much

data is needed for good statistics.

It was observed that the dynamic structure factor decays rapidly over time – in a matter of tens of picoseconds – so the density of data points in this short period of time is an important factor in determining the collective diffusion factor accurately. The above mentioned parameters were tested only with the Martini model and the results from tests were assumed to be applicable as such to the two other models. The Martini model was chosen for the tests because of its computational efficiency.

Sensitivity Analysis

The density fluctuations method was tested against different lengths of simulation data. Collective diffusion was calculated from separate simulations using the Martini model at 310 K and ranging from 10 to 200 ns. The results are presented as a function of simulation length in Fig. 5.2. The results clearly show that even with short simulations it is possible to gain accurate results and the method does not show a strong dependence on simulation length.

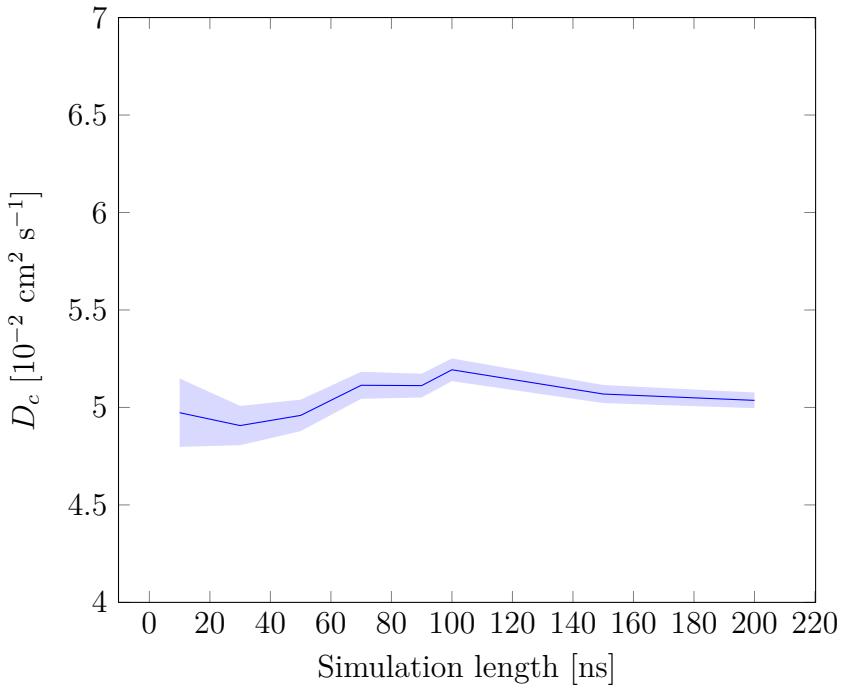


Figure 5.2: Collective diffusion for Martini model at 310 K as a function of simulation length.

The second simulation parameter that was tested was the frame output frequency in the trajectory file. This test allows us to examine the sensitivity with regard to data point density, which is important for the accuracy of the density-density autocorrelation function. First, one 100 ns simulation with 0.1 ps frame interval was run. Then from this simulation frames with sparser interval were chosen and

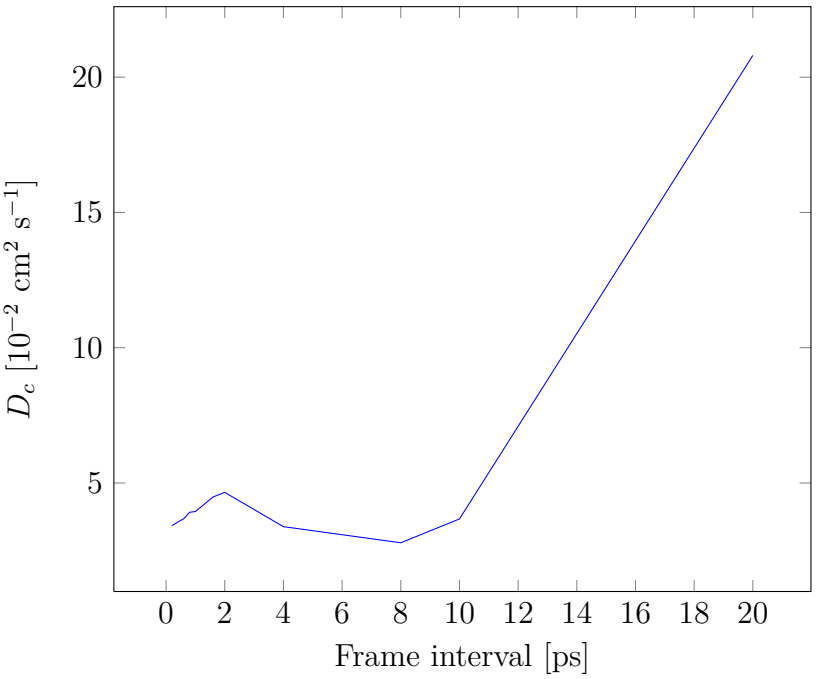


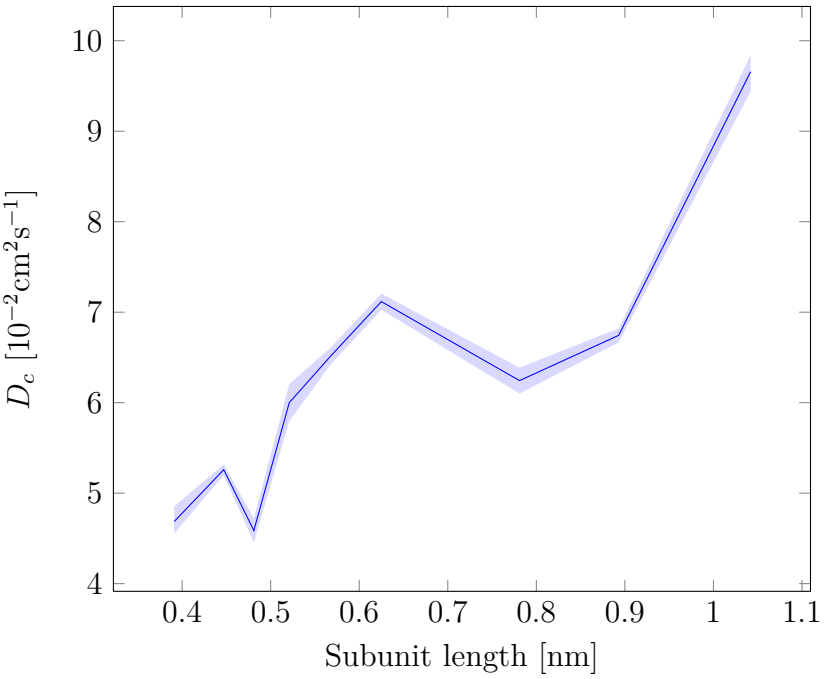
Figure 5.3: Collective diffusion for Martini model at 310 K as a function of frame interval

the analysis was run on the sparser set of frames. The results will not be statistically independent, but the effect of larger frame intervals becomes clear.

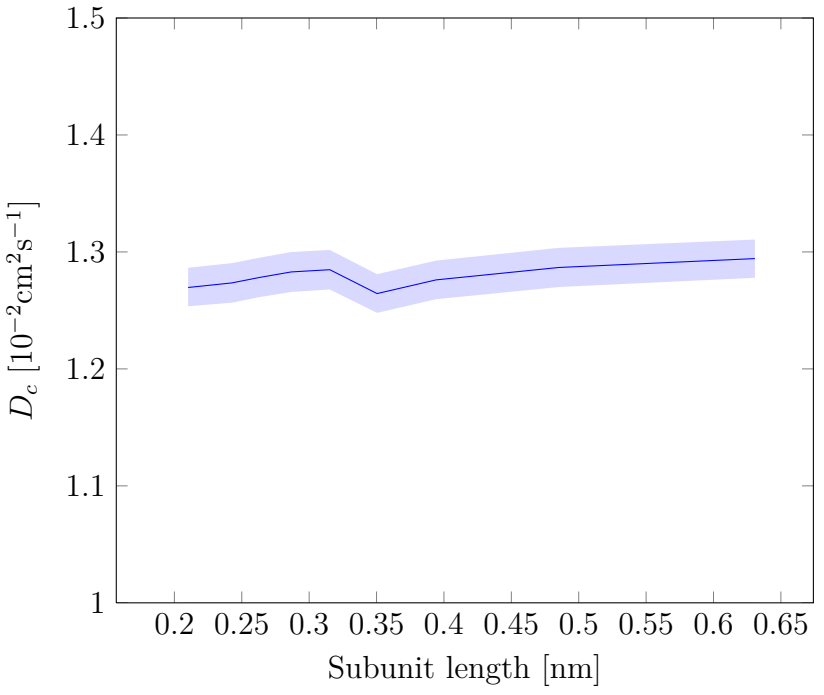
As indicated in Fig. 5.3 a simulation with output frequency of 1 ps is a good compromise between accuracy and efficiency. The dependence on frame interval is weak up to 10 ps, but beyond that a significant error starts to accumulate. Thus 1 ps frame interval was selected for the following final analyses.

As discussed in Chapter 3, when calculating the local density fluctuations, the simulation box needs to be divided into smaller subunits and the choice of subunit size is essential. In Fig. 5.4 the effect of subunit dimension is displayed with the smallest k-vector. The simulation box is 6.26 nm in length.

The Martini model shows a strong increasing linear trend with subunit length: collective diffusion doubles when subunit length grows 2.5 times. Around 0.47 nm there seems to be a kink which is interpreted to reflect the structure of the fluid. This interpretation is supported by the fact that the Martini Lennard-Jones –parameter is 0.47 nm, but since the gaps between the data points increase at lengths larger than 0.62 nm makes it too uncertain to draw further conclusions. Due to technical details in the analysis program, it is impossible to create a set of data points which would be closer together. In the following analyses the subunit length 0.47 nm is used because it matches the physical dimensions of the underlying particles. The polarizable Martini is assumed to behave similarly to Martini and the same choice of subunit length is made when analysing D_c for polarizable Martini. It is indeed worth bearing in mind that the choice of subunit length has a big impact on the



(a)



(b)

Figure 5.4: Collective diffusion (a) for CG and (b) for SPC as a function of subunit dimension at 300 K.

value of collective diffusion.

Interestingly, there is only a minimal dependence on the subunit length with the SPC model. A weak increasing linear trend can be observed with a dip from the general trend at 0.35 nm. The location of the first peak in the radial distribution function for SPC is known to be at 0.276 nm, which is not reflected in the figure. There seems to be no obvious explanation for the dip at 0.35 nm. For the results to be comparable with CG collective diffusion, the same unit length as above was used in SPC analyses.

Temperature-dependence

The temperature dependence of collective diffusion is presented in Figure 5.5 and in Table 5.3 for the first k-vector, and in Table 5.4 for the second k-vector. It can be observed that the ratio of the values remains fairly constant at different temperatures, which implies that the temperature-dependent behaviour is similar in all models. When comparing the values at different k-vectors, the strong k-vector-dependence is obvious, which is in odds with theoretical predictions. This raises doubts about the reliability of the chosen computation method or its implementation.

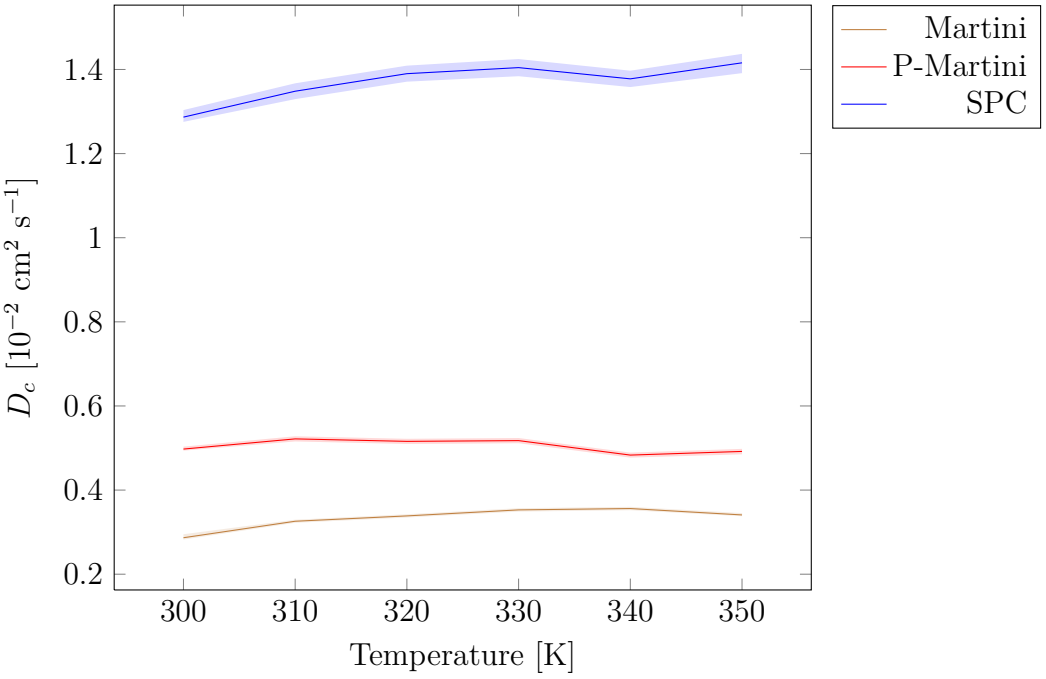


Figure 5.5: Collective diffusion as a function of temperature in the range 300–350 K.

The ratio of fine-grained collective diffusion to coarse-grained is surprisingly similar to the one measured for self-diffusion. The ratios differ only by a factor 1.5, which indicates that the time scales for these dynamic properties are similar.

Table 5.3: Collective diffusion for the first k-vector ($k = 1$) and the ratio between fine-grained and coarse-grained values. The values are in simulation time and coarse grained values have been transformed into their fine-grained representatives.

Temperature [K]	D_c [10^{-2} cm^2 s^{-1}]			Speed up factors	
	SPC	Martini	P-Martini	Martini/SPC	P-Martini/SPC
300	1.297	4.541	7.711	3.502	5.947
310	1.366	5.193	8.151	3.803	5.969
320	1.416	5.429	8.124	3.835	5.739
330	1.438	5.694	8.222	3.959	5.716
340	1.420	5.783	7.742	4.073	5.453
350	1.468	5.573	7.953	3.796	5.417

Table 5.4: Collective diffusion for the second k-vector ($k = 2$) and the ratio between fine-grained and coarse-grained values. The values are in simulation time and coarse grained values have been transformed into their fine-grained representatives.

Temperature [K]	D_c [10^{-2} cm^2 s^{-1}]			Speed up factors	
	SPC	Martini	P-Martini	Martini/SPC	P-Martini/SPC
300	0.088	0.516	0.878	5.844	9.950
310	0.102	0.544	0.937	5.335	9.189
320	0.110	0.693	0.983	6.315	8.965
330	0.123	0.744	1.020	6.072	8.328
340	0.126	0.775	1.049	6.157	8.334
350	0.131	0.808	1.082	6.172	8.265

There is a lot of difficulty in interpreting the results because of the strong k-vector dependence, which is clearly visible when comparing the values in Tables 5.4 and 5.3. Unfortunately no experimental data was found for comparison and the speed up factors with respect to real water cannot be obtained.

5.3 Viscosities

Two types of viscosities have been examined: shear viscosity and kinematic viscosity. Although they are closely related they complete each other in forming the picture of momentum transfer in the fluid.

The periodic perturbations method has been used and a thorough analysis of the method applied to atomistic water models has been done by Song *et al.* [28]. Thus the focus here is on the coarse grained models. Song *et al.* discovered that when using the periodic perturbations method, the shear viscosity depends on the chosen acceleration amplitude and indeed the same results is found in this thesis as illustrated in Fig 5.6. The apparent viscosity grows with decreasing acceleration amplitude, but at very small values for acceleration amplitude the signal-to-noise -ratio becomes poor and the results become unreliable.

The choice of acceleration amplitude is crucial for later simulations, where the temperature dependence is studied. Following Song *et al.*, the acceleration amplitude for SPC simulations was chosen to be 0.005 nm ps^{-2} and for CG simulations 0.002 nm ps^{-2}.

If the shear rate, which depends on the acceleration amplitude, is too high the system is too far from the equilibrium and thus is not able to relax. According to Hess, the system will be able to relax, if the inverse shear rate is longer than the typical relaxation time, such as the rotational relaxation time, 10 ps [37]. In the simulations in this thesis, the inverse relaxation times are between 21–35 ps in SPC simulations, 62–103 ps in polarizable Martini, and 95–170 ps in Martini simulations.

The temperature dependence of the models shows a correct trend in comparison to experimental data, but the polarizable Martini as well as the SPC model have consistently too low values. The Martini water displays reasonably accurate values in the temperature interval 300 to 320 K.

In contrast to diffusion factors, no geometric scaling is considered in the computation of viscosity, which results in dismissing the effective speed up factor of CG simulations. This is because the definition of shear viscosity in the periodic perturbation method does not take time explicitly into account nor does the definition of shear viscosity provide a simple method to determine this factor. The estimation of the geometric scaling factor arising from the CG process is extremely difficult.

Table 5.6 presents the ratios of viscosities, which can be interpreted as the speed up factors of the combined effect of the CG process and force field parametrization.

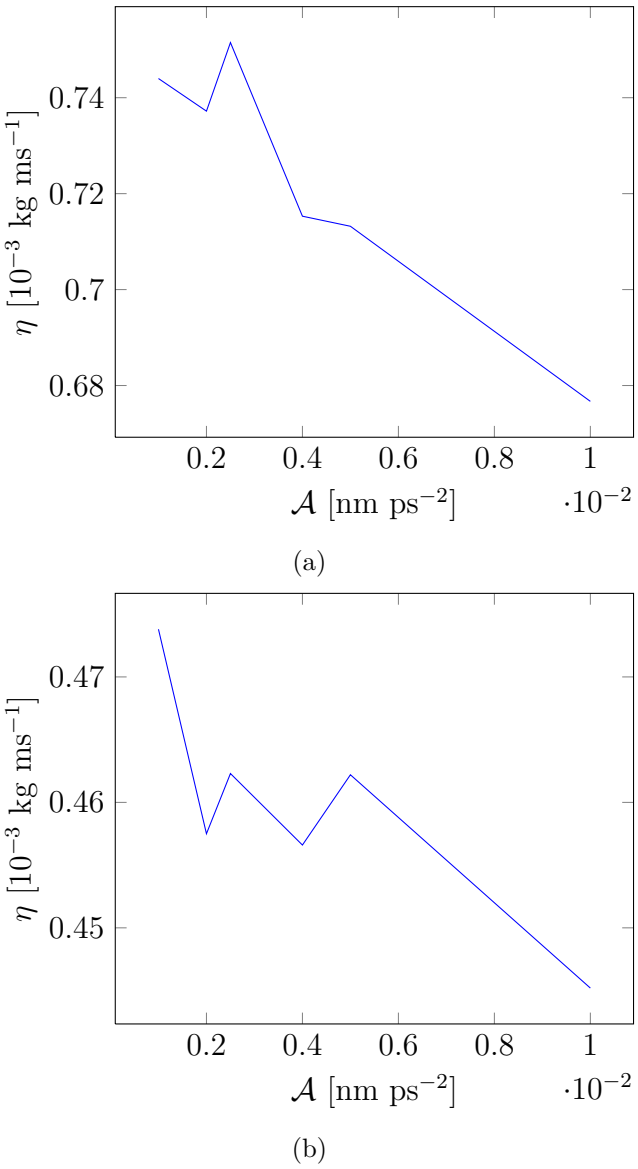


Figure 5.6: The shear viscosity as a function of acceleration amplitude (a) Martini model (b) polarizable Martini water as determined with the periodic perturbations method at 310 K.

The ratios are calculated by dividing the respective values in Table 5.5. There is a weak temperature dependence on the ratio and overall it is clear that the speed up factor here differs from the speed up factor in self-diffusion albeit the difference is a mere factor of 2.

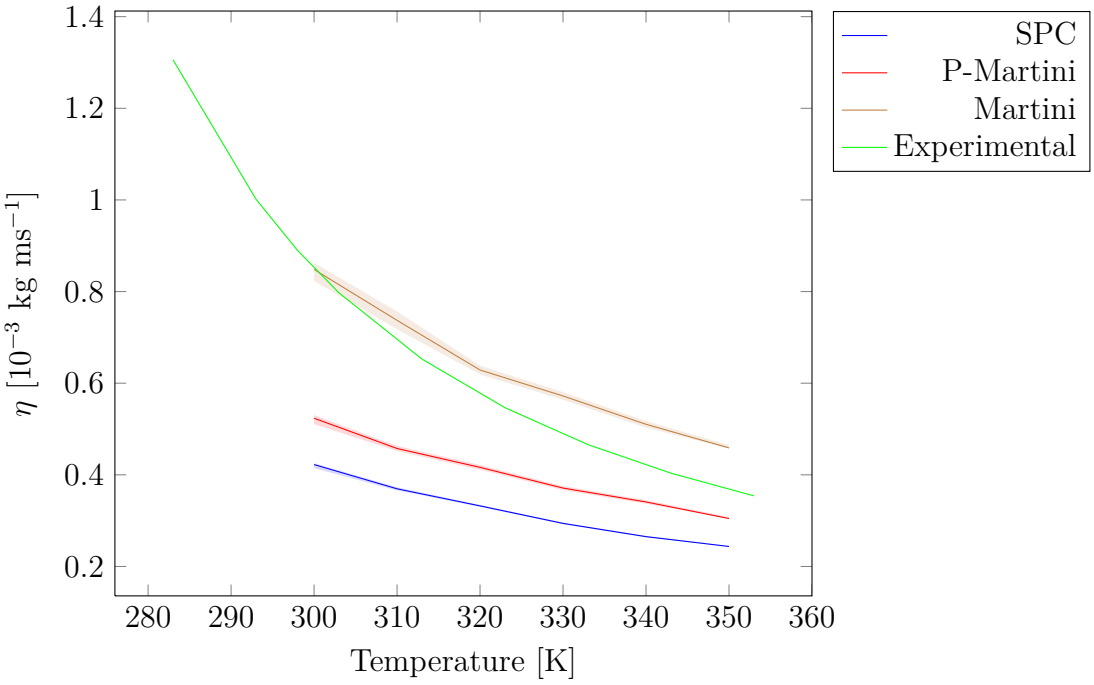


Figure 5.7: The shear viscosities as determined with the periodic perturbations method in the temperature range 300–350 K. Experimental data from [20]

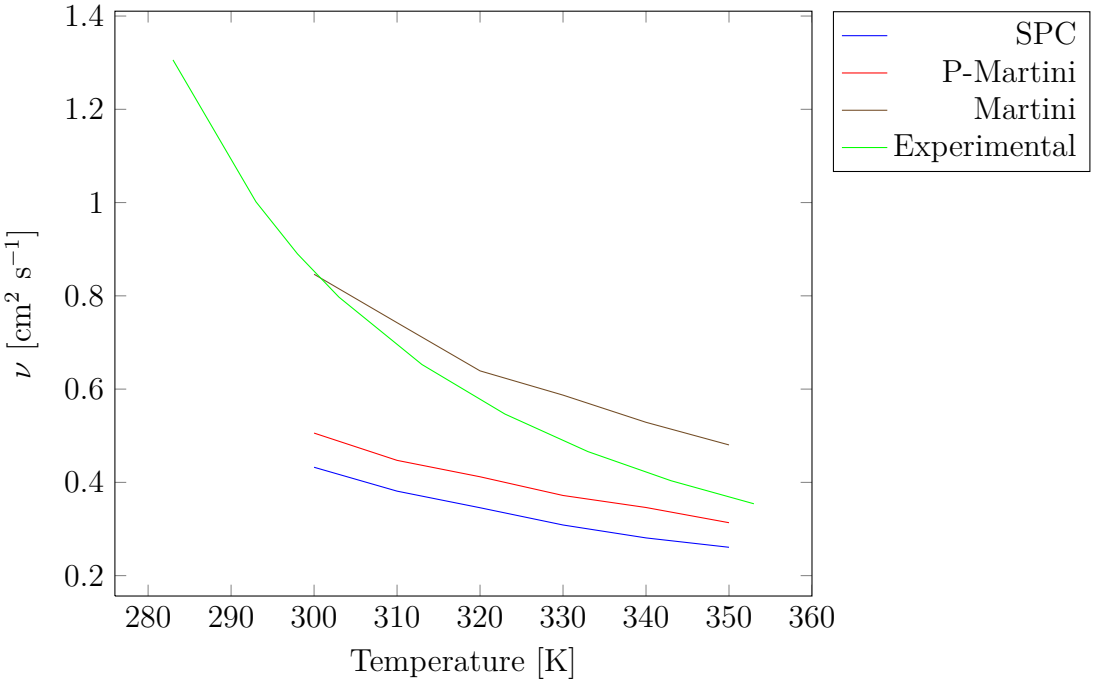


Figure 5.8: The kinematic viscosities as determined with the periodic perturbations method in the temperature range 300–350 K. Experimental data from [20]

Table 5.5: The shear viscosities as determined with the periodic perturbations method in the temperature range 300–350 K. Experimental data from [20]

	η [10^{-3} kg ms^{-1}]			
T [K]	SPC	P-Martini	Martini	Exp
300	0.422 ± 0.003	0.524 ± 0.007	0.848 ± 0.015	0.854
310	0.370 ± 0.003	0.458 ± 0.006	0.737 ± 0.02	0.693
320	0.332 ± 0.002	0.416 ± 0.006	0.629 ± 0.01	0.577
330	0.294 ± 0.002	0.371 ± 0.005	0.572 ± 0.009	0.489
340	0.265 ± 0.002	0.341 ± 0.004	0.510 ± 0.008	0.422
350	0.244 ± 0.002	0.305 ± 0.004	0.459 ± 0.007	0.369

Table 5.6: The ratios of shear viscosity which represent the speed up factors.

T [K]	SPC/Exp	Martini/Exp	P-Martini/Exp	Martini/SPC	P-Martini/SPC
300	0.495	0.994	0.613	2.008	1.240
310	0.533	1.063	0.660	1.995	1.238
320	0.576	1.090	0.722	1.893	1.253
330	0.601	1.170	0.759	1.946	1.262
340	0.629	1.210	0.808	1.925	1.286
350	0.661	1.245	0.827	1.883	1.251

Table 5.7: The kinematic viscosities as determined with the periodic perturbations method in the temperature range 300–350 K. Experimental data from [20]

	ν [cm^2 s^{-1}]			
T [K]	SPC	P-Martini	Martini	Exp
300	0.432	0.506	0.846	0.857
310	0.381	0.447	0.742	0.698
320	0.346	0.412	0.639	0.583
330	0.309	0.372	0.587	0.497
340	0.281	0.346	0.529	0.430
350	0.261	0.314	0.480	0.378

Furthermore the dynamic viscosity was determined from the periodic perturbation simulations. The kinematic viscosity ν scales the shear viscosity with density and thus gives a description of the diffusivity of momentum in the system. Due to the fact that all three models give consistent results with density, the picture given by kinematic viscosity is similar to shear viscosity. The difference between SPC and polarizable Martini becomes smaller where as the difference between the afore mentioned and Martini force field remains the same or is slightly increased. Martini model gives accurate values at room temperature, but overestimates the kinematic viscosity as the temperature increases. Since the ratios of the kinematic viscosities are essentially the same as for shear viscosity, no separate table is presented.

5.4 Schmidt Number

One useful tool to characterize the nature of fluid flows is the dimensionless Schmidt number, which describes the ratio of momentum diffusivity to mass diffusivity. In other words it can be used to tell whether the fluid is gas- or liquid-like. Gases have low Schmidt numbers where as thick liquids have high Schmidt numbers. In Fig. 5.9 the Schmidt number has been calculated in the temperature interval 300–350 K. For experimental results the data was gathered from [20, 44]. The values for correct temperatures were intrapolated using cubic splines from the existing data points.

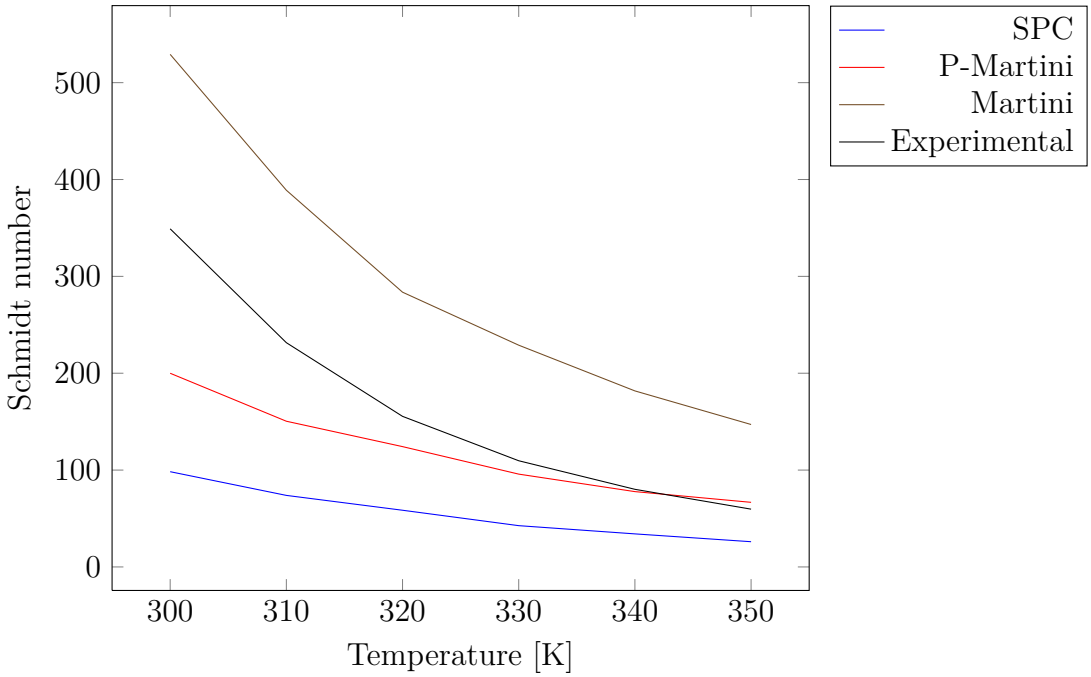


Figure 5.9: The Schmidt number in the temperature range 300–350 K. Experimental data interpolated from [20, 44]

The polarizable Martini and SPC force fields have weaker response to changes in temperature and the values are lower than the experimental ones. At around 300 K the Schmidt number of ethanol is 135.5 and we can see that polarizable Martini and SPC in particular have dynamic characteristics closer to ethanol than water. The Martini shows a correct trend in temperature-dependence, but the values are 1.5 to 2.4 times larger than the experimental values. Martini water can be considered thicker than real water, and SPC and polarizable Martini can be considered thinner than real water.

6. CONCLUSIONS

It is well known that the time evolution in CG simulations is faster than in atomistic simulations, but detailed understanding of the this has remained unclear. The dynamic properties of SPC, Martini, and polarizable Martini water models have been examined and compared in order to elucidate the question of time scales in fine-grained and coarse-grained simulations. The chosen dynamic properties have been self-diffusion, collective diffusion, shear viscosity and kinematic viscosity. Self-diffusion and collective diffusion present the picture of mass transfer and collective properties of the liquid whereas shear viscosity and kinematic viscosity measure momentum transfer. The main research questions in this thesis have been

1. How do the dynamic properties of coarse grained water models correspond to real water and on the other hand to atomistic water models?

2. Are there differences in the time scales of different dynamic properties?

In all of the studied properties the temperature-dependence of the models shows similar trend, but the actual values are distinctly different. The polarizable Martini model exaggerates the most the values of self-diffusion and with shear viscosity none of the models can mimic the correct temperature-dependence of real water. Experimental values depend much more strongly on temperature than the models can portray. Interestingly the models show only a minimal temperature-dependence with collective diffusion. Composite characteristic properties such as the Schmidt number reveal that the Martini model has the correct trend with temperature whereas the polarizable Martini has the most accurate values with experimental results.

In this study the observed speed up factors in coarse-grained simulations are in line with the semi-quantitative nature of the Martini coarse-grained force field. That is to say, when comparing the values for different dynamic properites the speed up factors are constant up to a factor of two. However, it must be noted that it is difficult to draw conclusions from the comparisons, because they have been made with figures from slightly different origin. The self-diffusion speed up has been calculated comparing D_s of the assumed fine-grained particles in simulation time to experimental values. For collective diffusion the speed up was gained by comparing the assumed fine-grained paritcles in simulation time with values from atomistic simulations with SPC, whereas the speed up for shear viscosity has been

drawn from comparing the values for coarse-grained particles in simulation time with experiments. From a theoretical point of view the effect of coarse-graining process on shear viscosity is still unclear and thus we are left with values, which only reflect the combined effect of the coarse-graining process and the force field parametrization. As a limiting factor, the results in this thesis hinge on the assumption of uncorrelated movement of the particles, which is not reasonable in many cases. For strongly interacting liquids, like water, this can be the case.

The results found in this study are also in line with earlier findings in the field. The time scale for pore formation and the self-diffusion coefficients for different lipids vary, but are close to the commonly held speed up factor of four [6]. The results support the current understanding, that the time scales in coarse-grained simulations are non-unique and there is no precise way of interpreting them. Even when studying the same simple isotropic system but using different dynamic measures, it is possible to observe different speed ups.

In further research it could be beneficial to look at the interplay of different time scales in complex simulations. What happens when phenomena depending on momentum transfer, which seems to have a slower speed up, are combined with phenomena depending on pore formation or collective diffusion, which have faster speed up factors.

BIBLIOGRAPHY

[1] **Benioff, M., Lazowska, E., et al.**, Computational science: ensuring America's competitiveness. *Report to the President, President's Information Technology Advisory Committee, Washington, DC*. 2005.

[2] **Hepolehto, K. and Vierros, E.**, CSC Vuosikertomus. 2011.

[3] **Shaw, D., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R., Eastwood, M., Bank, J., Jumper, J., Salmon, J., Shan, Y., et al.**, Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002): 341–346. 2010.

[4] **Shinoda, W., DeVane, R., and Klein, M.**, Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Molecular Simulation*, 33(1-2): 27–36. 2007.

[5] **H.J.C. Berendsen, J.P.M. Postma, W.v.G. and Hermans, J.**, Interaction models for water in relation to protein hydration. In B. Pullman, ed., *Intermolecular Forces*, pp. 331–342, D. Reidel Publishing Company, Dordrecht. 1981.

[6] **Marrink, S., de Vries, A., and Mark, A.**, Coarse Grained Model for Semiquantitative Lipid Simulations. *Journal of Physical Chemistry B*, 108(2): 750–760. 2004.

[7] **Venturoli, M., Maddalena Sperotto, M., Kranenburg, M., and Smit, B.**, Mesoscopic models of biological membranes. *Physics Reports*, 437(1): 1–54. 2006.

[8] **Kubo, R.**, The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29: 255. 1966.

[9] **Hansen, J. and McDonald, I.**, *Theory of simple liquids*. Academic press, 2^{nd} edition. 2000.

[10] **Onsager, L.**, Reciprocal Relations in Irreversible Processes. I. *Physical Review*, 37(4): 405. 1931.

[11] **Tuckerman, M.**, G25.2651: Statistical Mechanics, Notes for Lecture 21. `http://www.nyu.edu/classes/tuckerman/stat.mech/lectures/lecture_21/lecture_21.html`, retrieved: 28.06.2011.

[12] **Allen, M. and Tildesley, D.**, *Computer simulation of liquids*, volume 18. Oxford University Press. 1989.

[13] **Vattulainen, I. and Mouritsen, O.**, Diffusion in Membranes. In P. Heitjans and J. Kärger, eds., *Diffusion in Condensed Matter*, pp. 471–509, Springer, Berlin. 2005.

[14] **Gomer, R.**, Diffusion of adsorbates on metal surfaces. *Reports on Progress in Physics*, 53: 917. 1990.

[15] **Van Hove, L.**, Correlations in space and time and Born approximation scattering in systems of interacting particles. *Physical Review*, 95(1): 249. 1954.

[16] **Guillot, B.**, A reappraisal of what we have learnt during three decades of computer simulations on water. *Journal of Molecular Liquids*, 101(1-3): 219–260. 2002.

[17] **Wu, Y., Tepper, H.L., and Voth, G.A.**, Flexible simple point-charge water model with improved liquid-state properties. *The Journal of Chemical Physics*, 124(2): 024503. 2006.

[18] **Ichikawa, K., Kameda, Y., Yamaguchi, T., Wakita, H., and Misawa, M.**, Neutron-diffraction investigation of the intramolecular structure of a water molecule in the liquid phase at high temperatures. *Molecular Physics*, 73(1): 79–86. 1991.

[19] **Badyal, Y.S., Saboungi, M.L., Price, D.L., Shastri, S.D., Haeffner, D.R., and Soper, A.K.**, Electron distribution in water. *The Journal of Chemical Physics*, 112(21): 9206–9208. 2000.

[20] **Weast, R.**, ed., *Handbook of Chemistry and Physics*. CRC, Cleveland. 1977.

[21] **Marsh, K.**, ed., *Recommended Reference Materials for the Realization of Physicochemical Properties*. Blackwell, Oxford. 1987.

[22] **van der Spoel, D., van Maaren, P., and Berendsen, H.**, A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *The Journal of Chemical Physics*, 108: 10220. 1998.

[23] **Errington, J. and Panagiotopoulos, A.**, A fixed point charge model for water optimized to the vapor-liquid coexistence properties. *The Journal of Physical Chemistry B*, 102(38): 7470–7475. 1998.

[24] **Marrink, S., Risselada, H., Yefimov, S., Tieleman, D., and de Vries, A.**, The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *Journal of Physical Chemistry B*, 111(27): 7812. 2007.

[25] **van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A., Apol, E., Meulenhoff, P., Tieleman, D., Sijbrers, A., Feenstra, K., van Drunen, R., and Berendsen, H.**, Gromacs User Manual version 4.5. 2010.

[26] **Baron, R., de Vries, A.H., Hünenberger, P.H., and van Gunsteren, W.F.**, Comparison of Atomic-Level and Coarse-Grained Models for Liquid Hydrocarbons from Molecular Dynamics Configurational Entropy Estimates. *The Journal of Physical Chemistry B*, 110(16): 8464–8473. 2006.

[27] **Baron, R., Trzesniak, D., de Vries, A.H., Elsener, A., Marrink, S.J., and van Gunsteren, W.F.**, Comparison of Thermodynamic Properties of Coarse-Grained and Atomic-Level Simulation Models. *ChemPhysChem*, 8(3): 452–461. 2007, ISSN 1439-7641.

[28] **Song, Y. and Dai, L.**, The shear viscosities of common water models by non-equilibrium molecular dynamics simulations. *Molecular Simulation*, 36(7-8): 560–567. 2010.

[29] **Bulavin, L.A., Lokotosh, T.V., and Malomuzh, N.P.**, Role of the collective self-diffusion in water and other liquids. *Journal of Molecular Liquids*, 137(1): 1–24. 2008.

[30] **Leach, A.**, *Molecular modelling: principles and applications*. Prentice Hall, Harlow. 2001.

[31] **Ewald, P.**, Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, 64: 253–287. 1921.

[32] **Onsager, L.**, Electric moments of molecules in liquids. *Journal of the American Chemical Society*, 58(8): 1486–1493. 1936.

[33] **Darden, T., York, D., and Pedersen, L.**, Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98: 10089. 1993.

[34] **Essmann, U., Perera, L., Berkowitz, M., Darden, T., Lee, H., and Pedersen, L.**, A smooth particle mesh Ewald method. *Journal of Chemical Physics*, 103(19): 8577–8593. 1995.

[35] **Hockney, R. and Eastwood, J.**, Computer simulation using particles. *McGraw-Hill Inc.* 1981.

[36] **Luty, B., Tironi, I., and van Gunsteren, W.**, Lattice-sum methods for calculating electrostatic interactions in molecular simulations. *The Journal of Chemical Physics*, 103: 3014. 1995.

[37] **Hess, B.**, Determining the shear viscosity of model liquids from molecular dynamics simulations. *The Journal of Chemical Physics*, 116: 209. 2002.

[38] **Bussi, G., Donadio, D., and Parrinello, M.**, Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126: 014101. 2007.

[39] **Berendsen, H., Postma, J., van Gunsteren, W., DiNola, A., and Haak, J.**, Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81: 3684. 1984.

[40] **Vattulainen, I.**, *Studies of Surface Diffusion Under Equilibrium and Non-Equilibrium Conditions*. Ph.D. thesis, University of Helsinki, Helsinki. 1997.

[41] **Palmer, B.**, Transverse-current autocorrelation-function calculations of the shear viscosity for molecular liquids. *Physical Review E*, 49(1): 359. 1994.

[42] **Ciccotti, G., Jacucci, G., and McDonald, I.**, "Thought-experiments" by molecular dynamics. *Journal of Statistical Physics*, 21(1): 1–22. 1979.

[43] **Evans, D. and Morriss, G.**, *Statistical mechanics of nonequilibrium liquids*. Cambridge University Press. 2008.

[44] **Holz, M., Heil, S.R., and Sacco, A.**, Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate ^{1}H NMR PFG measurements. *Phys. Chem. Chem. Phys.*, 2: 4740–4742. 2000.