Attila Vrabecz

Classical Simulation Aided Studies on the Structure of Liquids

Ph. D. Thesis

Supervisor:
Dr. Gergely Tóth
Associate Professor

Eötvös Loránd University, Faculty of Science

PhD School of Chemistry,
Director: Dr. György Inzelt Professor

Theoretical and Physical Chemistry, Structure-Research Program,
Director: Dr. Péter Surján Professor

2007

List of Acronyms

BGY	Born-Green-Yvon equation (or hierarchy)
HNC	Hypernetted Chain approximation
MC	Monte Carlo simulation
MD	Molecular Dynamics simulation
OZ	Ornstein-Zernike equation
PY	Percus-Yevick approximation
RFA	Rational Function Approximation
RMC	Reverse Monte Carlo simulation

Acknowledgements

Of all people who have contributed to the accomplishment of the work reported in this thesis, my wife Anikó Kende definitely deserves to be first in line: thank you for all your support and for loving me - I do not know what I would do without you. I would also like to take this opportunity to thank for the rest of my family for believing in me and supporting me at all times.

I would like to express my gratitude to my supervisor Gergely Tóth for his continuous support, guidance and encouragement as much as in my research as in all matters of life.

I would also like to acknowledge the contribution of Norbert Király to the second chapter.

I would thank for the opportunity provided by the Chemistry Ph.D. School for providing me an environment to conduct my research and granting me a scholarship which made this thesis possible.

Table of Contents

LIST OF ACRONYMS **2**

ACKNOWLEDGEMENTS **3**

TABLE OF CONTENTS **4**

INTRODUCTION **6**

1 THEORETICAL INTRODUCTION **7**

1.1 STRUCTURE OF DISORDERED SYSTEMS 7

1.1.1 Potential Function *8*

1.1.2 Pair Correlation Function *11*

1.1.3 Diffraction Experiments *12*

1.1.4 Structure Factor *13*

1.2 COMPUTER SIMULATIONS 14

1.2.1 Molecular Dynamics *15*

1.2.2 Monte Carlo Simulations *15*

1.3 INVERSE PROBLEM 16

1.3.1 Integral Equations *16*

1.3.2 Reverse Monte Carlo *18*

1.3.3 Iterative Refinement of model potentials *19*

2 PAIR POTENTIALS FROM DIFFRACTION DATA ON LIQUIDS: A NEURAL NETWORK SOLUTION **21**

2.1 INTRODUCTION 21

2.2 DETAILS OF THE CALCULATIONS 22

2.3 RESULTS AND DISCUSSIONS 27

2.4 CONCLUSIONS 33

3 SIMULATION OF BINARY HARD SPHERE SYSTEMS WITH 1:5 AND 1:10 SIZE RATIOS **34**

3.1 INTRODUCTION 34

3.2 DETAILS OF THE CALCULATIONS 37

3.3 RESULTS AND DISCUSSIONS 40

3.3.1 Quantitative Limits of the One-Component Hard-Sphere Model *40*

3.3.2 Comparison to Percus Yevick and Rational Function Approximation Pair Correlation Functions *43*

3.3.3 Contact Values *47*
3.4 CONCLUSIONS 50
4 PARAMETERIZATION OF COULOMB INTERACTION IN THREE-DIMENSIONAL PERIODIC SYSTEMS 52
4.1 INTRODUCTION 52
4.2 PARTITIONING OF THE POTENTIAL ENERGY IN PERIODIC SYSTEMS 55
4.3 DETAILS OF THE CALCULATIONS 58
4.4 RESULTS AND DISCUSSIONS 62
4.5 CONCLUSIONS 71
5 SUMMARY 73
BRIEF SUMMARY 75
ÖSSZEFOGLALÁS 76
BIBLIOGRAPHY 77
LIST OF TABLES 82
LIST OF FIGURES 83
APPENDIX 85
P-7 POLYNOMIAL 85
R-6/1 POLYNOMIAL 86
R-4/3 POLYNOMIAL 86
R-3/3 POLYNOMIAL 87
P-3 POLYNOMIAL 87

Introduction

The enormous growth in affordable computational capacity (according to Moore's law (Thompson and Parthasarathy 2006)) eased the spreading of computer simulations, which created a new branch of science called computer experiments. Models, which are represented by mathematical functions, replace real materials and computers provide the space of experiments.

Computer simulations are used in virtually all aspects of science (physics, chemistry, biology, materials science, etc.) and its use is expected to grow substantially in applied science (pharmaceutical industry, chemical technology) as well. Computer simulations are useful in substituting experiments, where the studied system allows for a simple description, which still gives correct results on the examined properties. Another use of computer simulations is the study of system response to extreme conditions (pressure, temperature, etc.), where experimental setup is not available. Computer simulations are similar to both theories and experiments, however at the same time also complement both: theories can be tested by comparing theoretical results with computer simulations, while a model can be evaluated by comparing experimental results and computer simulations.

This work concentrates on the structure of disordered systems with the help of classical simulations; the articles on which this work is based are not tightly related. The three main topics are the following. Chapter 2 presents the part of the study where the aim was to assess whether with the usage of neural networks it is possible to solve the inverse problem of statistical mechanics; the method is tested on commonly used simple pair-potentials. In Chapter 3 the aim was to simulate binary hard-sphere system with large size ratios as those systems are of great scientific interest and there was no comprehensive study especially regarding pair-correlation functions with large size ratios to be found in the literature. This work involved algorithmic developments to be able simulate those systems with reasonable run times. In Chapter 4 the aim was to develop an efficient parameterisation of the Coulomb-interaction for simulation uses.

Remark: all the program codes during this work were created from scratch except where it explicitly stated otherwise and during their development numerous technical, algorithmic and numeric problems were faced.

1 Theoretical Introduction

In the first part of this literature review I present an overview of the necessary theoretical concepts. In the following part short overview of computer simulations is given in the context of determining liquid structure. The last section presents the inverse problem of statistical mechanics and the conventional solutions.

1.1 Structure of Disordered Systems

There are two general structural functions in the liquid field: pair-correlation function (see 1.1.2) and structure factor (see 1.1.4). It is known in statistical mechanics since the article of Henderson (Henderson 1974), that there is a one to one correspondence between the interaction function and the structure of simple liquids. If a system can be described as particles interacting with classical pair potentials (see 1.1.1), then the pair correlation functions and the pair interactions uniquely determine each other at a given density and temperature in equilibrium. The only uncertainty is a constant in the pair potential, but it is usually set to ensure zero interaction at infinite distance. The theorem is proved several times (Henderson 1974; Hansen and McDonald 1986; Zwicker and Lovett 1990; Tóth and Baranyai 2000; Baranyai and Schiller 2003) and it is valid for multi component systems as well (Hansen and McDonald 1986; Tóth and Baranyai 2000). It is true also for complicated molecular systems as long as molecular distribution functions are used (Allen and Tildesley 1987). The theorem cannot be applied rigorously if site-site distribution (Allen and Tildesley 1987) functions are used, because the site-site representation can be obtained from the molecular distribution functions by integrating out variables (Hansen and McDonald 1986), in practice known molecular structures limit the uncertainty of the site-site representation.

The pair correlation function is the inverse Fourier-transform of the structure factor obtained by diffraction measurements. In case of multi-component systems, it is valid both for total and partial functions (see 1.1.2). Fourier-transformation is unique for continuous functions on both directions. Thus, the relationship of the pair potential and the pair correlation function can be extended to the structure factor. However, the structural functions are known only at discrete

values and the proof of the theorem cannot be held. If the functions are known in enough length and with sufficient resolution, the theorem is still applicable.

The theorem states the relationship, but it does not say how to determine the two functions. Computer simulations (see 1.2) provide an exact solution to get the pair correlation function for a known interaction (Allen and Tildesley 1987; Haberlandt et al. 1995; Frenkel and Smit 1996). For some extreme cases, like colloids, we can still use the approximate methods of the integral equation theories of liquids being popular at the time of very limited computer power (Hansen and McDonald 1986).

There is a lack of exact, robust and computationally widespread method in the opposite direction, where pair interactions are determined from known structural functions. The determination of the functions in this direction often referred as the inverse problem of statistical mechanics (see 1.3). The early attempts used the approximations of the integral equation theories (see 1.3.1) on liquids (Hansen and McDonald 1986). Later on, methods applying computer simulations appeared, most of these methods are iterative (see 1.3.3).

1.1.1 Potential Function

The most important part of a model is the interaction function (potential) of the particles. In an unordered system at a given temperature and in a given volume the attractive and repulsive forces result in a mutual interaction potential. The full potential energy (E_p) depends on the interactions between the different particles, on the distance between particles and their positions relative to each other; as well as on the number of interacting particles. The starting point of computer simulations is the description of particle interactions. These interactions can be derived from various sources: quantum-mechanical calculations, experiments and theories. To use quantum-mechanical precision throughout the whole simulation of a system is still not feasible because of the enormous computational demand.

1.1.1.1 Pair Potentials

To calculate the full potential of M molecules we write E_p as a sum. This sum consists of the energies of the possible *ij* pairs, *ijk* triplets, *ijkl* quadruplets, etc.

$$E_p = \sum_{i<j} \varphi_{ij} + \sum_{i<j<k} \Delta\varphi_{ijk} + \sum_{i<j<k<l} \Delta\varphi_{ijkl} + ... + \Delta\varphi_{123...M} \qquad \textbf{(1.1)}$$

φ_{ij} is the interaction energy of the *i*, *j* particle pair if we leave out of consideration the rest of the molecules. $\Delta\varphi_{ijk}$ is the *i*, *j* and *k* particle-triple's mutual energy increase. The $\Delta\varphi_{123...M}$ term comes from the mutual interaction of *M* molecules. The increase in potential energy belongs to the whole system and it is not contained in previous terms.

E_p is pairwise additive if all the terms are zero except for the first one. For low-density systems the interaction energy is often pairwise additive, mathematically

$$E_p = \sum_{i}^{M} \sum_{j>i}^{M} \varphi_{ij} \qquad \textbf{(1.2)}$$

However this assumption can cause errors at higher densities.

Assuming pairwise additivity one can get better results if the potential is derived from properties of the crystal of the material, because the density of the solid phase deviates less from the liquid phase. Fitting the different properties of the crystal (lattice constant, compressibility factor, etc.) to the theoretically optimal shape, the parameters of the potential can be derived (Momany et al. 1974). This way one can get the Huggins-Mayer functions (Larsen et al. 1973), which are good for alcali-halogenids; and the Williams potential functions (Jorgensen et al. 1984), which was developed for organic molecules. Depending on the purpose of the modelling intramolecular potentials of bending, stretching and torsion can be taken into account.

If simple interaction models can describe a system there is no need to use more complex interactions. Simple interaction models are for example hard cylinders or spheres, soft cylinders or spheres, Lennard-Jones particles, Buckingham (exp.-6) model, piecewise linear functions and 1-1 charged Coulomb particles.

1.1.1.2 Hard Sphere Potential

In this model particle shape is approximated by a hard sphere. It is also the simplest potential used in practice.

$$\varphi^{HS}(r) = \begin{cases} 0, r \geq d \\ \infty, r < d \end{cases} \qquad \textbf{(1.3)}$$

where *r* is the distance between the particles and *d* is the diameter of the spheres.

1.1.1.3 Lennard-Jones Potential

The Lennard-Jones potential could be used to describe the phase behaviour of pure noble gases and simple sphere-like molecules:

$$\varphi^{LJ}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad \textbf{(1.4)}$$

where ε and σ are the Lennard–Jones parameters and $\sigma = r_0\left(\frac{1}{2}\right)^{1/6}$.

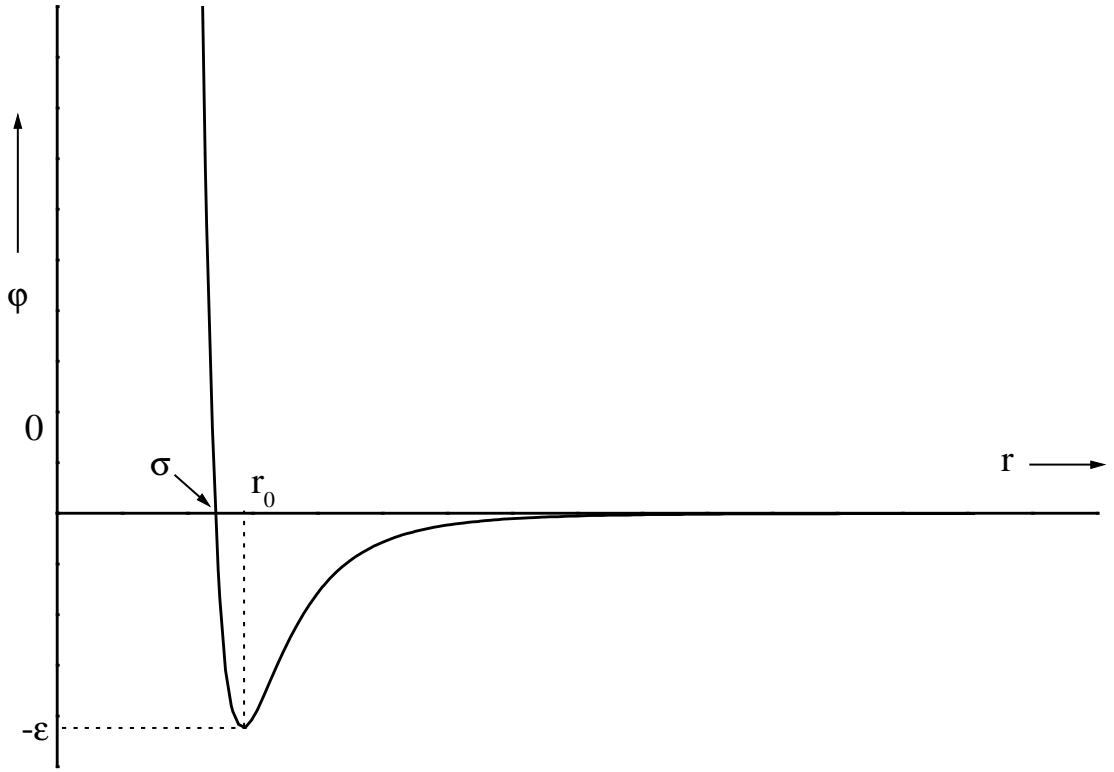


Figure 1.1 Typical Lennard-Jones (6, 12) potential

On Figure 1.1 σ denotes the point where the graph cuts the $\varphi = 0$ axle, and r_0 denotes the place minimum of the potential ($\varphi = -\varepsilon$). The φ potential energy results as the sum of the terms from the positive repulsive and negative attracting forces. The Lennard-Jones σ is the zero point of the repulsive part of the potential and the minimum is at $\sqrt[6]{2}\sigma$.

1.1.1.4 Morse Potential

$$\varphi^{M}(r) = D\{1-\exp[-A(r-r_0)]\}^2 - D \tag{1.5}$$

where r denotes the scalar distance of two particles, r_0 is the position of the minima of the potential, D and A are parameters.

1.1.1.5 Buckingham Potential

The formula for the modified Buckingham potential:

$$\varphi^{B}(r) = \frac{\varepsilon_B}{1-6/\alpha}\left\{\frac{6}{\alpha}\exp\left[\alpha\left(1-\frac{r}{r_B}\right)\right]-\left(\frac{r}{r_B}\right)^6\right\} \tag{1.6}$$

r denotes the scalar distance of two particles, r_B is the minima of the potential, ε_B and α are parameters.

1.1.2 Pair Correlation Function

To describe the microscopic structure of liquids the pair correlation function - $g(r)$, also called radial distribution function - used most of the time, which gives the mutual, pairwise distribution of the particles:

$$g(r) = \frac{\rho(r)}{\rho_0} \tag{1.7}$$

where $\rho(r)$ is the local number density from an arbitrary particle in r scalar distance, ρ_0 is the average number density of the system. $\rho_0 = N_{par}/V$, where N_{par} is the number of particles in the system and V is the system volume. The $g(r)$ function can be given in the knowledge of the coordinates of all particles, which constitute the system:

$$g(r) = \frac{n(r)}{4\pi\rho_0 dr} \tag{1.8}$$

where $n(r)$ is the number of particles from the arbitrarily chosen central particle in distance r within a dr infinitely thick spherical shell. The pair correlation function in distance r from the chosen particle reflects the relationship between the local and average number density. In disordered systems at long distances there is no correlation between the particle positions, so

$$\lim_{r\to\infty} g(r) = 1 \tag{1.9}$$

As the atoms have a well-defined volume

$$\lim_{r\to 0} g(r) = 0 \qquad \textbf{(1.10)}$$

In the case of more components partial pair correlation functions can be devised:

$$g_{\alpha\beta}(r) = \frac{\rho_{\alpha\beta}(r)}{\rho_\beta} = \frac{n_{\alpha\beta}(r)}{4\pi r^2 \rho_\beta \mathrm{d}r} \qquad \textbf{(1.11)}$$

So $g_{\alpha\beta}(r)$ denotes the probability of finding a type α particle within *r* distance by a type β one. Type α and type β particles can be the same as well. To have an unambiguous notation in the case of the component one of the components will be denoted with 1, while the second component will be denoted 2, but α and β will be used in the sums, where the indices can relate to both components. So by writing $g_{\alpha\beta}$ we refer to three functions: g_{11}, g_{22} and g_{12}=g_{21}.

We can also define multiple-body correlation functions similar to the pair correlation function. Complex structures can only be described unambiguously by the multiple-body correlation functions, however these are not accessible from experimental data, and they can only be calculated in the knowledge of the positions of all particles.

There are some frequently used transformed forms of the pair correlation function. Integrating $4\pi r^2\ \rho(r) = 4\pi r^2\ \rho^0\ g(r)$ from $r = 0$ it is possible to calculate the corresponding coordination number for the different coordination shells.

It is also important to note that in the case of pairwise additive potentials the pair correlation functions are sufficient to calculate thermodynamic properties, e.g. energy, pressure.

1.1.3 Diffraction Experiments

The phenomenon of diffraction is the change of direction of an incident radiation on matter attributed to the interaction between the radiation wave and the material. In a diffraction experiment the intensity of the scattered radiance is measured as a function of the scattering angle (2Θ). The scattering angle is the angle defined by the wave vectors of the incident and scattered wave.

About the static structure the elastic – no energy transfer – diffraction gives information. The angle dependency and intensity patterns depend on the structure of the material and the incident radiation's wavelength. The diffracted waves interfere with each other: in case of the same phase interference amplification (constructive interference), in case opposite phase interference extinction (destructive interference) happens. The diffraction picture is the appearance of reflections according to the constructive interference as a function of the diffraction angle. For

perfect structures (e.g. ideal crystals) at particular Θ angles perfect wave amplification can be expected (Figure 1.2). These are the so-called Bragg-angles, which satisfy the following equation:

$$n\lambda = 2d\sin\Theta \tag{1.12}$$

where $n\lambda$ is an integral number multiple of the wavelength, d is the distance of the planar grids. For not perfect structures the perfect wave amplification is impossible, sharp diffraction peaks cannot be observed. The intensity of the reflections carries information as well.

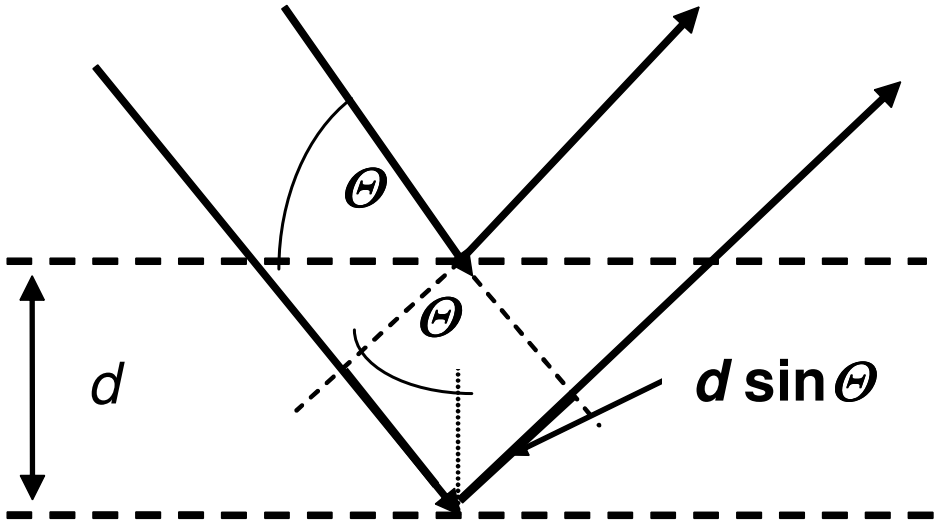


Figure 1.2 Diffraction, Bragg-angles

For particles in the 1-100 nm size interval (e.g. colloids) or in case of inhomogenities the coherent scattering phenomena has the information at small angles, which is the reason why small angle techniques are used for studying these systems (Ramsay and Lindner 1993).

1.1.4 Structure Factor

The different diffraction experiments (neutron, X-ray, laser) used for the determination of the structure of liquids and other disordered materials result in the so-called static structure function or structure factor.

$$S(\mathbf{q}) = \frac{1}{N}\sum_k \sum_l \exp[-iq(\mathbf{r}_l - \mathbf{r}_k)] \tag{1.13}$$

where $\mathbf{q}$, the difference of the incident and diffracted waves vectors is the diffraction variable, $\boldsymbol{r}_l$ and $\boldsymbol{r}_k$ are position vectors of the particles l and k respectively.

Based on statistical mechanical principles it can be proven that $S(q)$ and $g(r)$ are related:

$$S(\mathbf{q}) = 1 + \rho \int \exp(-i\mathbf{q}r)[g(r) - 1]dr \qquad \textbf{(1.14)}$$

It can be seen that the structure factor is the Fourier-transform of the pair correlation function. If the system is isotropic, then wave numbers can be used instead of the wave vectors: $q = |\mathbf{q}|$, where $q = (4\pi/\lambda)\sin\Theta$, is the diffraction angle, λ is the wavelength corresponding to the particle. Thus:

$$S(q) = 1 + 4\pi\rho_0 \int_{r=0}^{\infty} r^2 [g(r) - 1] \frac{\sin(qr)}{qr} dr \qquad \textbf{(1.15)}$$

Defining the relationship with the inverse transformation:

$$g(r) = 1 + \frac{1}{\rho(2\pi)^2} \int_0^{\infty} q^2 \frac{\sin(qr)}{qr} [S(q) - 1] dq \qquad \textbf{(1.16)}$$

It can be concluded that applying a Fourier-transformation on $S(q)$, which is obtained from a diffraction experiment, one can get $g(r)$ which can be interpreted in real space, however the calculation of the pair correlation function this way raises some issues. The first one is that because $S(q)$ is finite the $g(r)$ obtained with Fourier-transformation is accused of a decreasing cut-off error as r increases. The second issue is that systematic errors in the structure factor cause false oscillations before the true first maxima of $g(r)$.

1.2 Computer Simulations

Computers are used for simulations almost since their first appearance (Metropolis N et al. 1953; Alder and Wainwright 1956). For computational experiments on liquids popular methods are the classical mechanical ones, like the Metropolis Monte Carlo (MC) simulations and molecular dynamics (MD) (Allen and Tildesley 1987; Haberlandt et al. 1995; Frenkel and Smit 1996). However there are some hybrid techniques combining features from both which are not included in this thesis.

Most times computations become feasible only if several numerical tricks and well thought-out algorithms are used. These tricks and the details of the algorithms are not detailed here only if it is relevant to this current work. Descriptions of the most widely used algorithms and numeric tricks can be found in the works of (Allen and Tildesley 1987; Rapaport 1995; Frenkel and Smit 1996).

1.2.1 Molecular Dynamics

The molecular dynamics method follows the behaviour of the system in time. The particles have positions, and - according to the temperature and to the corresponding Maxwell-Boltzmann velocity-distribution - momentum coordinates. The particles are moved at each time step according to the calculated forces, derived from the assumed interaction potential. As the calculation of the forces is time-consuming the algorithms favoured to be used only calculate the forces once per time-step. Such algorithms are e.g. the one-correction step predictor-corrector, the Verlet-Störmer and velocity Verlet methods. The new positions and momentums are obtained with integrating the Newton equation of motions at every time step. Enough steps done to reach equilibrium, after that thermodynamical properties can be obtained as a time-average. Molecular dynamics is deterministic.

1.2.2 Monte Carlo Simulations

Monte Carlo (MC) methods are a class of computational algorithms for simulating various systems. It is a stochastic way of solving a determinate problem transformed into an appropriate probabilistic formulation.

The basic outline of the MC method in this context is simple:

1. Try to displace a randomly chosen particle
2. Calculate the energy associated with the new configuration
3. If it is less than the previous one, accept the displacement
4. If it is not less accept it with $\exp(-\Delta U/kT)$ probability

It is proven that this algorithm ensures that the obtained configurations explore the phase-space appropriately. By selecting the average displacements' mean value one can influence the acceptance ratio. While it does not influence the result, it is best when optimised to be around 0.5, which should result in faster exploration of phase space.

One has to pay attention to the quality of the random number generator, as it is quite central to MC simulations. A 'bad' pseudo-random number generator can influence the results to a great extent especially in small systems.

The development of codes for MC simulations tends to be simpler than the corresponding MD one. The resulting code tends to run faster in the sense that calculating the same property for a given accuracy with an MC and an MD simulation, MC simulation takes a shorter amount of time

on usual computers used nowadays. This efficiency inspired many hybrid methods, which are not detailed here. It must be also noted that the MD method is much more suitable for vector/multiprocessor processing which seems to be the way of future.

As MC method does not use kinetics one cannot obtain kinetics-related data by performing MC simulations. This restricts the usage of the method, but when it can be used it might provide efficiency with both processing and development time because of the earlier mentioned reasons.

1.3 Inverse Problem

The inverse problem of statistical mechanics is to obtain the potential function from the pair correlation function or other experimentally measurable data like the structure factor.

1.3.1 Integral Equations

There are several theories for the inversion based on integral equations.

1.3.1.1 Born-Green-Yvon Equation

Let us have a system consisting of N particles. Let us denote the pair correlation function as $g_N^{(2)}(\mathbf{r}_1,\mathbf{r}_2)$. $g_N^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$ denotes the triplet correlation function, where the index superscript in parentheses relates to the rank of the distribution and the subscript index denotes the number of particles in the system. The Born-Green-Yvon equation (BGY) is the following:

$$kT\nabla_1 \ln g_N^{(2)}(\mathbf{r}_1,\mathbf{r}_2) = \nabla_1 \varphi(\mathbf{r}_1,\mathbf{r}_2) + \rho \int dr_3 \nabla_1 \varphi(\mathbf{r}_1,\mathbf{r}_3)\left(\frac{g_N^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)}{g_N^{(2)}(\mathbf{r}_1,\mathbf{r}_2)} - g_N^{(2)}(\mathbf{r}_1,\mathbf{r}_3) \right) \qquad \textbf{(1.17)}$$

if the interaction is pairwise additive.

The BGY equation is a formal connection between $g_N^{(2)}(\mathbf{r}_1,\mathbf{r}_2)$, $g_N^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$ and $\varphi(\mathbf{r}_1,\mathbf{r}_2)$. If any of them is known the third one can be determined. The issue is that only the pair correlation function can be devised from experimental data, one cannot know anything about the triplet correlation function. For solving the Equation (1.17) one has to apply some closure. One of these closures is the Kirkwood superposition approximation:

$$g_N^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3) = g_N^{(2)}(\mathbf{r}_1,\mathbf{r}_2) g_N^{(2)}(\mathbf{r}_1,\mathbf{r}_3) g_N^{(2)}(\mathbf{r}_2,\mathbf{r}_3) \qquad \textbf{(1.18)}$$

This relation implies that the triplet correlation of the particles is equal to the product of the pair-correlations. This means that each pair of three particle correlates with each other as if they

would be independent from the third one. It is worthwhile to note that this is a crude approximation and it is only true if the system is very dilute.

Solving the BGY equation we get a non-linear integral-differential equation for the pair correlation function – pair-potential connection. Equations developed later on generally surpassed the precision of the results (thermodynamical function estimation), but most of them started from this basic relation.

1.3.1.2 Ornstein-Zernike Equation

This equation has a greater practical significance. The definition of the total, direct and indirect correlation has to be introduced. The total correlation between two particles consists of the direct correlation between the two particles, but there is also a correlation between the relative positions of the two particles transmissioned indirectly by a third atom. This physical understanding is represented in the Ornstein-Zernike (OZ) equation:

$$h(\mathbf{r}_1,\mathbf{r}_2) = c(\mathbf{r}_1,\mathbf{r}_2) + \int d\mathbf{r}_3 \rho^{(1)}(\mathbf{r}_3) c(\mathbf{r}_1,\mathbf{r}_3) h(\mathbf{r}_2,\mathbf{r}_3). \quad \textbf{(1.19)}$$

This relationship can be looked at as the definition of the direct correlation function $c(r)=c(\mathbf{r}_1,\mathbf{r}_2)$, because the total correlation function $h(r)=h(\mathbf{r}_1,\mathbf{r}_2)$ can be expressed with the help of $g^{(2)}(r)=g(\mathbf{r}_1,\mathbf{r}_2)$ the following way:

$$h(r) = g^{(2)}(r) - 1 \quad \textbf{(1.20)}$$

From this it can be concluded that the OZ equation is expressing $g(r)$ with $c(r)$, so the problem will be to determine $c(r)$.

1.3.1.3 Percus-Yevick and Hypernetted Chain Approximation

The Hypernetted Chain (HNC) and Percus-Yevick (PY) theories are approximative methods to solve the OZ equation.

According to the PY approximation:

$$c(r) = [1-\exp(\beta\varphi(r))]g(r) \quad \textbf{(1.21)}$$

where $\beta = \frac{1}{kT}$ and k is the Boltzmann constant.

According to the other (HNC) important approximation:

$$c(r) = h(r)-\beta\varphi(r)-\ln g(r) \quad \textbf{(1.22)}$$

Substituting both equations to the OZ equation a closed expression can be derived in both cases for $g^{(2)}(r)$ and $\varphi(r)$.

The PY solution is a good approximation for systems which can be described by a short distance, and only repulsive interaction, while the HNC is better for models with long distance attractive potential (Hansen and McDonald 1986).

The HNC approximation can be improved with introducing a so-called bridge-function, $B(r)$ (Rosenfeld and Aschroft 1979). This bridge-function can be derived from the direct correlation function with applying a cluster series technique (not described here) and it denotes the elementary class of the clusters. Using the bridge function the direct correlation function takes the following form:

$$c(r) = h(r)-\beta\varphi(r)-\ln g(r) + B(r). \quad \textbf{(1.23)}$$

In HNC the $B(r) = 0$ approximation is used, if $B(r)$ is substituted with an appropriate reference system's known bridge function we get the RHNC (Reference HNC) formula. If the chosen reference system is the hard sphere liquid then the method is called MHNC (Modified HNC).

$$c(r) = h(r)-\beta\varphi(r)-\ln g(r) + B_0(r, \eta) \quad \textbf{(1.24)}$$

where η is the so called volume fraction.

In theory the previously discussed integral-differential equations are invertible and they result in the effective pair-potential. If we reorganise and study the HNC and PY approximations:

$$\varphi_{HNC}(r) = kT\,(h(r) - c(r) - \ln g(r)) \quad \textbf{(1.25)}$$

$$\varphi_{PY}(r) = kT \ln\left(1 - \frac{c(r)}{g(r)}\right) \quad \textbf{(1.26)}$$

We can come to the conclusion that to be able to solve the approximations (and get the potential) in both cases the knowledge of $c(r)$, $g(r)$ and $h(r)$ is required. For solving the inverse problem based on this idea more methods have been developed but the results are far from the expectations (Johnson et al. 1964).

1.3.2 Reverse Monte Carlo

The Reverse Monte Carlo (RMC) method was developed in 1988 by (McGreevy and Pusztai 1988). Basically it is quite similar to the MC method. The difference is that taking the sample taking is not based on energy in accordance with the equilibrium Boltzmann-distribution, rather on the χ^2, which is calculated from the difference between the configuration's calculated pair

correlation function ($g(r)$) and the experimental pair correlation function. Thus the RMC method does not use interaction functions.

The fundamentals of the process are the following:

From a given configuration of the particles a new one is derived with random placement(s) and χ^2 is calculated according to this.

For the old configuration:

$$\chi^2 = \sum_{i=1}^{n} \frac{(g_k(r_i) - g_s(r_i))^2}{\sigma_k^2(r_i)} \quad \textbf{(1.27)}$$

For the new configuration:

$$\chi'^2 = \sum_{i=1}^{n} \frac{(g_k(r_i) - g'_s(r_i))^2}{\sigma_k^2(r_i)} \quad \textbf{(1.28)}$$

where n is the number of measurement points, the k index relates to the experimental values, while s stands for simulation. σ_k is the assumed error of the experiment, while the prime denotes the new configuration.

If $\chi'^2 < \chi^2$, then the new configuration is accepted, whenever $\chi'^2 > \chi^2$ the configuration is accepted with according a normal probability distribution σ_k standard deviation. The process is continued, while χ^2 oscillates around a given value. Taking the sample cannot be solely done based on the difference in $g(r)$–s, but on the experimentally measurable structure factor ($S(q)$) or on any other property which is experimentally measurable and also can be calculated from configuration of the particles. In these cases obviously in Equations (1.27) and (1.28) $S(q)$ or the other corresponding functions should be used instead of $g(r)$.

1.3.3 Iterative Refinement of model potentials

Simulations are performed with a guess of the pair potential and the potential is modified on the comparison of experimental and simulation data. The procedure is iteratively repeated as long as the simulations reproduce the experimental data satisfactorily.

The first method was suggested by (Schommers 1973): he applied different methods (e.g. WeeksChandler-Andersen approximations, BGY equation, HNC-, PY theory) to study the connection of the effective pair-potential and the pair correlation function. Later on (Reatto 1982; Reatto 1988) suggested to use MHNC (as this is the best approximation in theory). Solving the inversion of MHNC he got an initial potential with which he performed MC simulation and with

the result of the simulation he corrected the original potential. He continued the iteration till he reached a fixed point: the potential did not change. Further ideas and improvements were published by (Lyubartsev and Laaksoonen 1995) and (Soper 1995). There are methods also from the last years: (Rutledge 2001), (Tóth 2001; Tóth 2003), (Almarza and Lomba 2003; Almarza and Lomba 2004), (Bourasseau et al. 2003) and (Wilding 2003). Most of these methods determine the interaction potential from pair correlation functions that limits their applicability, because many of the experimental structure factors can be transformed directly to pair correlation functions only with large uncertainty. Few methods start directly from the structure factor: a new method of (Soper 2001), the method of (Tóth 2001; Tóth 2003), of (Almarza and Lomba 2004) and of (Bourasseau et al. 2003). In references (Tóth 2001; Bourasseau et al. 2003; Tóth 2003; Soper 2005) it is possible to incorporate other experimental data into the process and to use total structure factors. Soper's method was rather popular in the last years. The method of Almarza and Lomba was used also to determine triplet interactions (Russ et al. 2005). A different approach was used in one of our previous studies. The Born-Green-Yvon hierarchy (BGY, 1.3.1.1) makes it possible to determine the pair potential exactly, if the pair correlation function and the triplet-correlation function of a system are known. They generated three-dimensional particle configurations using the reverse Monte Carlo (RMC) method (McGreevy and Pusztai 1988) on the basis of experimental data. They calculated the triplet- and the pair correlation functions of systems, and we applied a discrete version of the Born-Green-Yvon hierarchy to derive the pair-potentials (Tóth and Baranyai 1999; Tóth and Baranyai 2000). Unfortunately, the results were only approximate and they think the unphysical sampling of the reverse MC method was the main cause (Tóth and Baranyai 1997).

2 Pair Potentials from Diffraction Data on Liquids: a Neural Network Solution

2.1 Introduction

The determination of a pair-potential with an iterative method (see 1.3.3) requires at least one long Monte Carlo or molecular dynamics like simulation or several shorter ones and the process needs an initial guess of the pair interaction. Therefore, we searched for a 'quick and dirty' method. The success of the artificial neural networks in the solution of biological and computational problems was encouraging. The method is powerful in the approximation of non-linear functions. There is an excessive literature of the method and there are a large number of variants. Without going into details, the basic idea is sketched in Figure 2.1. The main components are the input layer, the neurons in the hidden layer and the output layer. Each channel of the input is connected to each neuron of the hidden layer. The input data are multiplied with weights; there is a separate weight for each input channel – neuron connection. The input of a neuron is the sum of the input data multiplied with the weights. A neuron is a mathematical function, e.g. tangent- or sine-hyperbolic, that gives an output as answer on the input. Each neuron is connected with each channel of the output, and there is a separate weight for each of these connections. The output is the sum of the weighted neuron answers at a channel. This structure is the most general one for neural networks, but one can find many variants, e.g. with more layer of neurons and differently defined network connections. There are two steps in the application of a simple neural network. At first, known set of inputs and outputs are used and the weights of the model are optimised to get a maximal reproduction of the desired outputs. This process is called training. In the second step, for known inputs we can get the unknown outputs simply by feeding the input into the trained neural network. This scheme of application is the most popular one, but there are many other strategies, like the use of Kohonen network without training.

Neural networks were used several times in chemical physics. There were applications connected to ab initio potential-energy surfaces and to the representation of classical force fields (Gassner et al. 1998; Cho et al. 2002; Hu et al. 2003; Raff et al. 2005). Neural networks can help in kinetic

modelling (Shenvi et al. 2004), in estimating vibration spectra in solutions (Lilichenko and Kelley 2001), in the evaluation of powder diffraction spectra (Habershon et al. 2004) and in the estimation of solubility of different species (Kiss et al. 2000; Panek et al. 2005).

We would like to show our neural network results on the solution of the inverse problem. The investigation was purely a numerical experiment, we produced a large number of structure factor – pair potential pairs by molecular dynamic simulation. We trained different neural networks with a part of our data, and we tested their power in the solution of the inverse theorem. The input data of the neural network were the structure factors and the output data were the pair potentials (Figure 2.1). We should state here, that this method is fully empirical, it is only a non-linear approximation of a function at different inputs and it has no rigorous connection to physics.

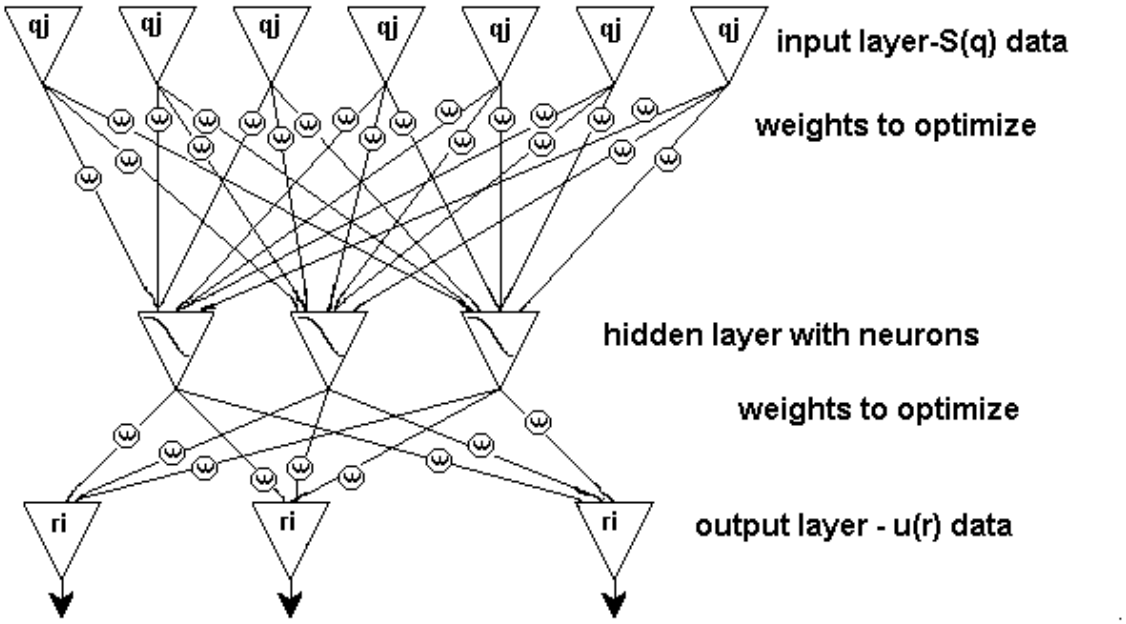


Figure 2.1 Schematic representation of an artificial neural network

2.2 Details of the Calculations

The structure factor – pair potential database was created with molecular dynamics simulations. The simulations were performed in the canonical ensemble, velocity-Verlet integration of time was applied and the temperature was regulated with Nose-Hoover thermostat. The systems contained one component and 1,000 particles were in the periodic cubic boxes. Three types of

interactions were applied: Lennard-Jones (1.4), Morse (1.5) and modified Buckingham (1.6) potentials (Allen and Tildesley 1987; Rapaport 1995; Frenkel and Smit 1996).

The potential at the minima are $-\varepsilon$, $-D$ and $-\varepsilon_B$. The factors in the exponentials (A and α) influence the hardness of the respective pair potential. In the case of the Buckingham potential, there is a maximum at small distances (r_{max}). If $r<r_{max}$, the modified Buckingham potential is usually set to infinity. The position of r_{max} depends on the α parameter. We chose r_B, α, the density, the temperature and the initial configurations in a way to avoid these distances in the simulations.

The potential parameters, the density and the temperature of the interactions were chosen randomly from given intervals. The intervals are summarised in Table 2.1. The second half of the table contains information on the realisation of the parameters in reduced units. The reduced units of the Lennard-Jones potential are general in chemical physics. Therefore, we translated our Morse and Buckingham potentials to a comparable unit by applying the approximation, that the unit reduced length was $r_0/\sqrt[6]{2}$ or $r_B/\sqrt[6]{2}$. We did not accept all pair potential – structure factor pairs provided by the simulations with the random parameters. We omitted the sets, where the systems were supposed to be solid ones, or where we suspected liquid-gas coexistence in the systems. The pair correlation functions and the structure factor-s were calculated over 50,000 time steps ($\Delta t=10^{-15}$s) after the equilibration of the systems. The final database contained pair potential–structure factor pairs of 1100 Lennard-Jones, 550 Morse and 550 Buckingham systems.

The neural network calculations were done by a program code written by us and by the NeuroSolutions program package (http://www.nd.com). Some preliminary calculations were performed with small data sets, where 90 or 600 structure factor (input) - pair potential (output) pairs were used as training set and the results were tested on 10 or 100 further data pairs. In the production runs 2000 pairs (1000 Lennard-Jones, 500 Morse and 500 Buckingham) were used as training set and 200 pairs (100 Lennard-Jones, 50 Morse and 50 Buckingham) as independent test one. The densities of the systems were further input for the neural networks. The temperatures of the systems were left out explicitly, because each pair potential (output) was rescaled into kT units. Many different calculations were performed to get an optimal solution. We tested possible network structures, different number of neurons, the effect of the data range, the normalisations in the input and the output, cut-off effects and different type of weight-training methods.

We focused on the simple multilayer models, because we did not expect more than a non-linear approximation of functions from the method. Networks with one, two and three hidden layers were tested. We obtained slightly better reproduction of the training data with more hidden layers than with a single one, but tests on the rest of the data were not convincing. We think the effect of 'over-training' was more emphasised for two and more hidden layers. It means, the weights were optimised for the training set perfectly, but the weights were very specific for the training set and did not provide a general solution of the problem. Modular networks were tried too, where the input data acted on several parallel networks which results were recombined to a final solution. The performance of modular networks was more limited than that of the simple multilayer models. The number of the neurons/hidden layer was optimised empirically. Sets of weight trainings (20-50 trials with 10,000-30,000 circles of weight updates) were performed with different number of neurons. In the case of the single hidden layer model, the best performance was achieved with 60 neurons on the test data set. The reproduction of the training set improved with higher number of neurons, but it seemed to result over-training. The weights were optimised with the static back-propagation algorithm (Jesús and Hagan 2007). An update of the weights was performed taking into account the whole training set. We tried with updates after each data set separately, but the convergence was not stable.

	Lennard-Jones	Morse	Buckingham
Interval of σ, r_0 or r_B (Å)	3.0-4.0	3.7-4.3	3.0-4.3
Interval of ε, D or ε_B (J/mol)	1-2000	1-2000	1-2000
Interval of for temperature (K)	50-150	50-150	40-200
Interval of ρ ($Å^{-3}$)	0.005-0.03	0.005-0.03	0.005-0.03
Interval of A or α		3-12	12-200
Minimal and maximal reduced temperatures in the selected systems	0.722-103	0.722-278	0.722-236
Minimal and maximal reduced densities in the selected systems	0.144-0.844	0.190-0.816	0.109-0.874

Table 2.1 Intervals for the randomly generated systems. Minima and maxima of the reduced quantities in the selected systems. See text for the reduced units.

The input functions of the networks were the structure factors and the densities. The output functions were the pair potentials in *kT* units. Theoretically, both functions were disposable in discrete form within a wide range with a very good resolution, because the *S(q)*-s were calculated with molecular dynamics and the *u(r)*-s were analytical inputs there. Practically, we truncated the functions at both low and high independent variables, and we used the data only at large and equidistant grid points. We left out the initial part of the pair potentials, because extra small pair distances never occurred at our low and medium temperatures. The pair potentials were truncated also at large *r* distances, where the interactions were close to zero. We applied the same [2.7, 14.9]Å distance range for all of the potentials, because it is not easy to use data with different ranges in neural networks. We used a grid of 0.2Å to reduce the number of the output channels. Finally, we got 62 output points. The structure factors were truncated over q=12Å^{-1} where *S(q)* was close to unity in our systems. The lower limit was q=1.6Å^{-1}, because structure factors determined with simulations may have anomalous behaviour (McGreevy and Pusztai 1988; Tóth 2003) at low q. Furthermore, the [1.6, 12]Å^{-1} range is usually available in the experiments that can be important in the later application of our method. Experimental data are measured at different q values than we applied, but they can be converted to our equidistant grid with interpolation methods. The truncation of the input data caused a loss of information, but it is not drastic in the case of simple liquids as it was shown by (Pusztai et al. 1995). The grid in the q space was 0.2Å^{-1} resulting 53 input channels for *S(q)* and one extra input channel for ρ.

The initial weights were randomly distributed in the (-0.5;0.5) range. Each input channel was normalised separately. Normalisation was not applied for the output channels, but a specific merit function was defined. The densities of the systems and the size of the particles ($\approx\sigma$) were different. It meant that the pair potential at the beginning of the used interval could be rather different, from a few to several hundred *kT*-s. An enormous repulsive part of *u(r)* had no relevant effect on the structure factor at our temperatures. In any optimisation process, the standard merit function is the error weighted square difference of the measured and calculated output data. It is valid for the back-propagation algorithm, too. In our case, an error of the pair potential (desired output data in neural networks) had no sense, because they were the fixed input data in the molecular dynamics simulations. We did not want to over-emphasise the initial and in the reality not used part of the potential, so after some trials we defined the merit function as:

$$< M >= \frac{1}{N_S \cdot N_M} \sum_{j=1}^{N_S} \sum_{i=1}^{N_M} \frac{\left(u_{ij}^{REAL} - u_{ij}^{CALC}\right)^2}{W_{ij}^2} \quad \text{, where}$$

$$W_{ij} = 1, \text{ if } u_{ij}^{REAL} \leq 1, \qquad \textbf{(2.1)}$$

$$W_{ij} = u_{ij}^{REAL}, \text{ if } 1 < u_{ij}^{REAL} \leq U_{cut}, \text{ and}$$

$$(1/W_{ij})^2 = 0, \text{ if } U_{cut} < u_{ij}^{REAL}$$

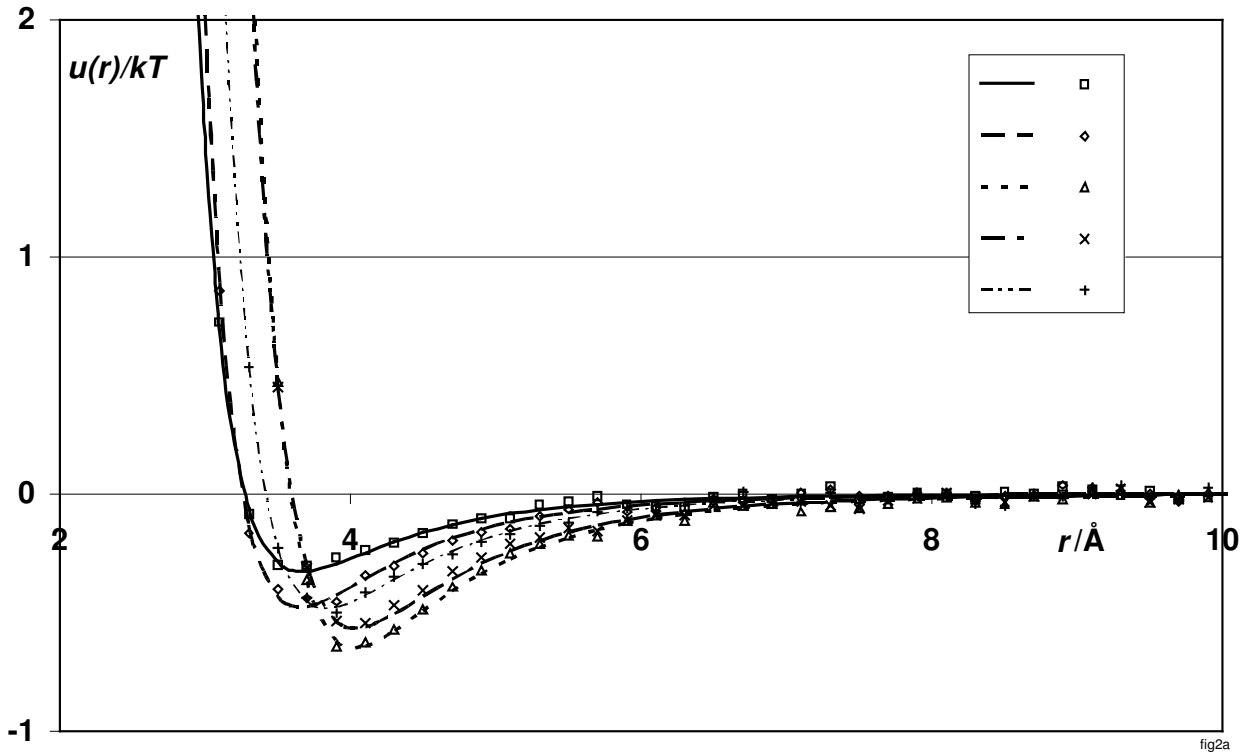
where u_{ij} was the tabulated pair potential at a given r_i pair distance for the j-th data set, N_M was the number of the used r_i distances, N_S was the number of data sets and U_{cut} was a potential cut-off in kT units.

The introduction of this merit function improved the performance of the networks significantly. A potential point was omitted from the calculation of the merit function, if it was larger than the cut-off. Calculations were performed with cut-offs of 10, 7.5, 5, 2.5 kT energies and with an infinite cut-off. The average, minimal and maximal merit functions were calculated for the training and the test data sets not only with the applied cut-off in the training, but also with the other cut-offs and the results were compared to each other. A high cut-off or its omission worsened the reproduction of the pair potentials in the important range at the given temperatures. The low U_{cut}=2.5 kT let the potential to be unphysical over 2.5 kT, while the r-range of 2.5 kT was involved in some collisions. 5 and 7.5 seemed to be suitable compromises. The merit function was calculated in this way to qualify the reproduction during the training and the test and in the optimisation process of the weights with the back-propagation algorithm. The merit function will not be used in the production calculations, where the unknown interaction will be determined for an input of a real experimental $S(q)$. We mention here, that the accuracy or the error of the input structure factors did not appear in our process. The training (and test) sets were calculated with simulations, where the uncertainty of $S(q)$ for a given T, ρ, N and pair potential was very small for simulation length applied by us. It could be easily estimated, because the uncertainty of a computational $S(q)$ is connected to the fluctuation in the simulation (Tóth and Baranyai 1997). In spite of its explicit omission from the training and test processes, the networks maybe treat this uncertainty in a mean field way. If the generalisation feature of the network works, the explicitly omitted input error would appear in the weights, because the weights were optimised on a large data set.

2.3 Results and Discussions

The data in this section were calculated with an input layer - single hidden layer – output layer network. The hidden layer consisted of 60 neurons with tangent hyperbolic functions. The multiplier was uniform in the tangent hyperbolic functions. The merit function of Equation (2.1) was used and the potential points were not used over 7.5 kT energy. The training set consisted of 2000 structure factor – pair-potential pairs. The performance of the method was tested on 200 pairs of functions. The neural networks were trained also to the different type of potentials separately: to the 1000 Lennard-Jones structure factor – pair-potential pairs, to the 500 Morse ones and to the 500 Buckingham ones. Here the test sets consisted of 100, 50 and 50 pairs, respectively. There was a training with the combination of the Lennard-Jones and Morse data, too. We performed about 100 trainings with different random starting weights for each case. The weights were updated maximum 30 thousand times in one calculation.

The performance of a network with given weights was measured with the merit function of Equation (2.1). The merit functions of the best-trained networks are summarised in Table 2.2. One can see, that the individual trainings to the different type of potentials resulted better reproduction than the training to all type of data simultaneously. The neural networks could be adapted to the Lennard-Jones potentials with the best performance. The reproduction of the Morse potential was slightly worse. We got the worst reproduction for the Buckingham potentials. It was valid for both individual trainings to the different type of potentials and the case, where all type of pairs were used in the training. The difference in the best and worst reproduction was about two magnitudes in the merit functions calculated on one-one pair potential. The 5 best established pair potentials of the training to all pairs are shown for the training set in Figure 2.2a and for the test set in Figure 2.2b. The best training and test systems were all Lennard-Jones ones. The reproduction was excellent here. In contrary, there are large differences in the 5 worst potentials shown for the training set in Figure 2.2a and for the test set in Figure 2.2b. There were two Morse and three Buckingham systems within the 5 worst training ones, and there were two Lennard-Jones, one Morse and two Buckingham systems within the 5 worst test ones. There were discrepancies around the minima of the potentials and the potentials obtained were not smooth at large r values.



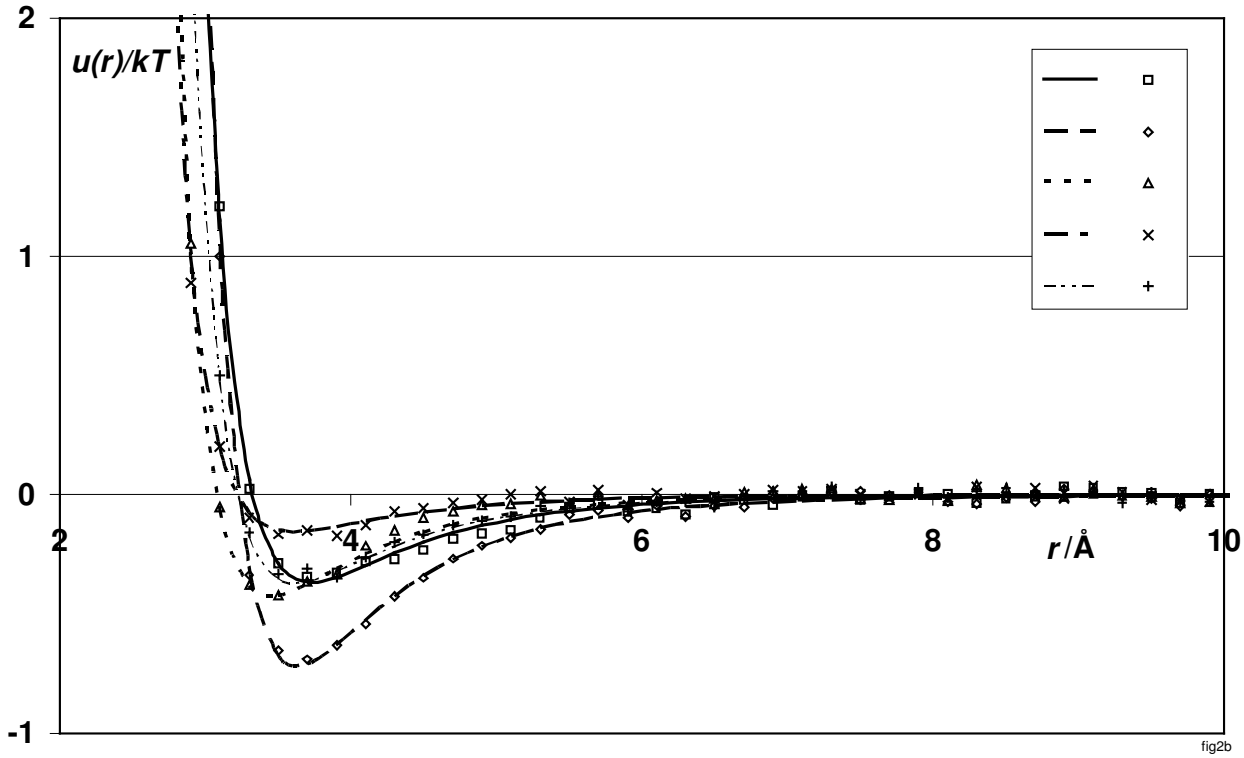
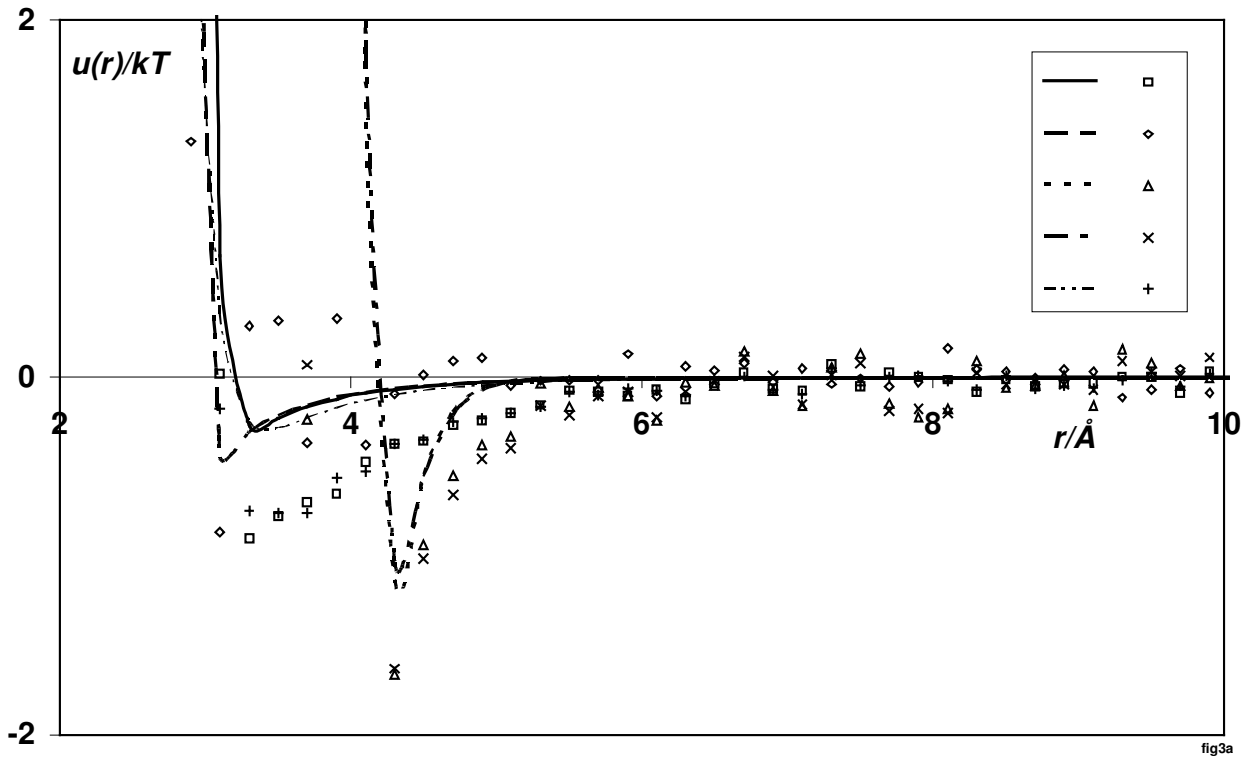


Figure 2.2 Five pair potentials reproduced best for the training (a) and test (b) sets. lines: original data, markers: output data



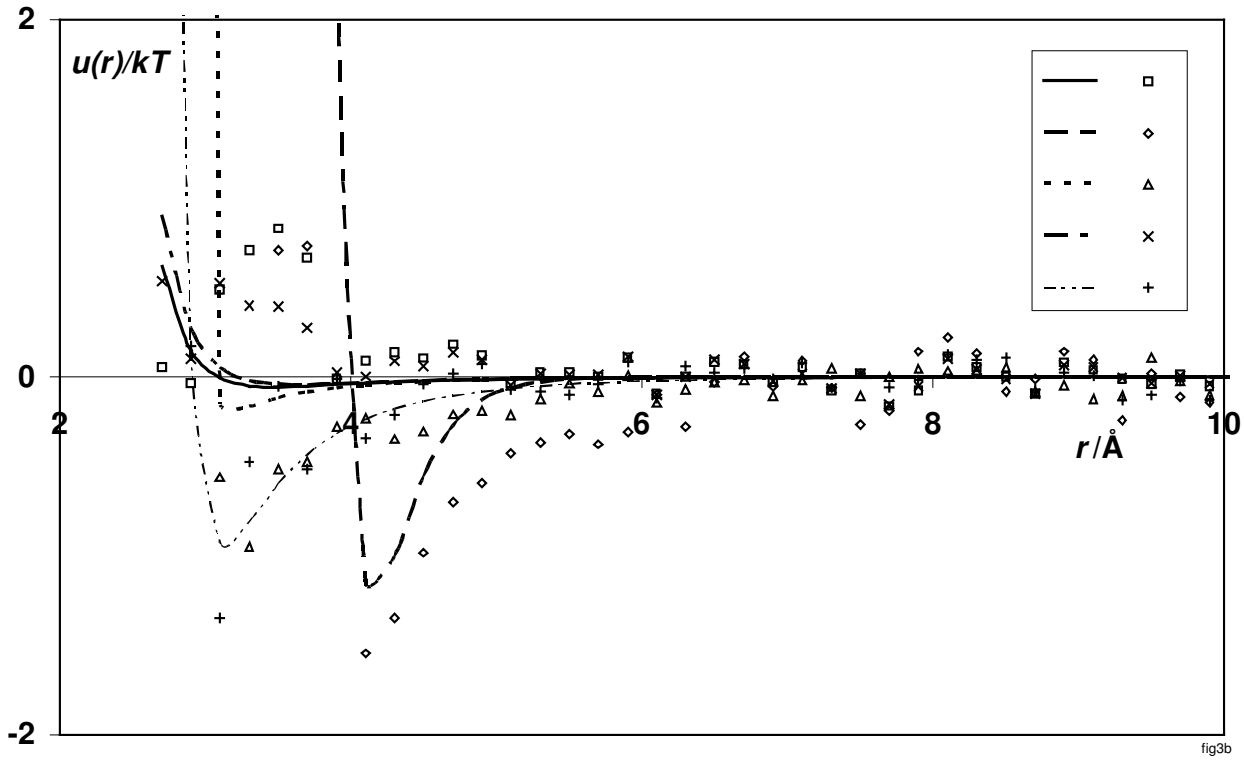


Figure 2.3 Five pair potentials reproduced worst for the training (a) and test (b) sets. lines: original data, markers: output data

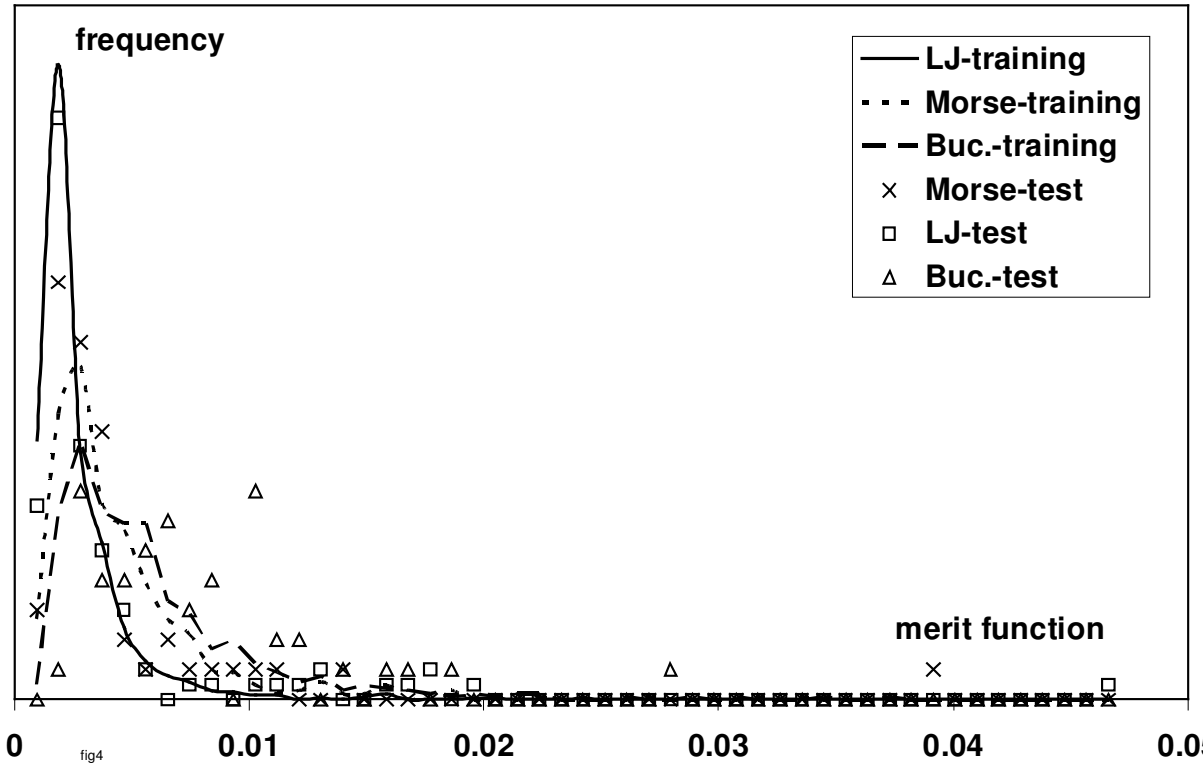


Figure 2.4 Distributions of the merit function for the training to all type of potentials simultaneously.

The worst reproductions seemed to be discouraging, but we checked the distribution of the individual merit functions for the potentials and we got rather asymmetric ones. The merit-function distributions calculated for the training and test sets of the three types of potentials for the training to the all data simultaneously are shown in Figure 2.4. One can see that the reproduction for the majority of the potentials was satisfactory and only a few exemplars were reproduced weakly and outstandingly. Depending on the training set and the test set we could define an efficiency measure for the method. We defined a percentage wherein the reproduction was qualitatively satisfactory. On a rather arbitrary way, we found that 0.004-0.005 merit function meant a satisfactory reproduction of a pair potential. The percentage of the data sets below 0.005 is shown in Table 2.2. The trends of these percentages were similar to the ones discussed for the merit function. As it was probable, the percentages were higher for the training sets than for the test sets. The differences were small meaning only slight over-training. Our neural networks usually kept their generalisation behaviour. Furthermore, the trainings to multiple types of potentials enhanced the reproduction of the Morse ones. Here the percentages in

the test potentials were higher than in the training ones. On the other hand, the Buckingham potentials could not be reproduced satisfactorily and the presence of these potentials in the training set decreased the efficiency drastically. We present the worst reproduced five data from the 90% best established potentials in the training and test sets of Lennard-Jones potentials. The potentials are qualitatively good in Figure 2.5. It means an application of the trained network with an input structure factor results as reliable potentials as those ones in Figure 2.5 with percentages shown in Table. Of course, the performance will depend on the similarity of the system to our Lennard-Jones, Morse and Buckingham systems.

training set	merit function of	training data				test data			
		min	<*M*>	Max	% of *M*<0.005	min	<*M*>	max	% of *M*<0.005
All	All	0.00036	0.00352	0.03674	79.8	0.00051	0.00468	0.04660	70.7
	Lennard-Jones	0.00036	0.00225	0.02397	93.2	0.00051	0.00373	0.04660	85.6
	Morse	0.00057	0.00416	0.03299	73.9	0.00077	0.00401	0.03863	80.90
	Buc.	0.00064	0.00543	0.03674	59.2	0.00134	0.00730	0.02791	32.4
Lennard-Jones & Morse	Lennard-Jones&Morse	0.00021	0.00242	0.03573	89.1	0.00025	0.00287	0.04041	87.8
	Lennard-Jones	0.00021	0.00189	0.02598	93.1	0.00025	0.00265	0.04041	86.8
	Morse	0.00028	0.00348	0.03573	81.0	0.00079	0.00331	0.02206	90.1
Lennard-Jones	Lennard-Jones	0.00011	0.00065	0.00575	99.8	0.00016	0.00102	0.00857	98.1
Morse	Morse	0.00018	0.00110	0.00566	99.6	0.00041	0.00249	0.02252	92.2
Buc.	Buc.	0.00033	0.00223	0.01233	93.9	0.00044	0.01022	0.12083	51.4

Table 2.2 Merit functions of the networks trained best. The minima and maxima correspond to the best and worst reproduced pair potentials.

We checked the dependence of the merit function on the reduced temperature or density. We could not detect any correlation; the reduced parameters of the best and the worst function corresponded well to a random choice from the ensemble of the systems. Furthermore, we could not detect any correlation between the merit functions and the analytic parameters of the potentials.

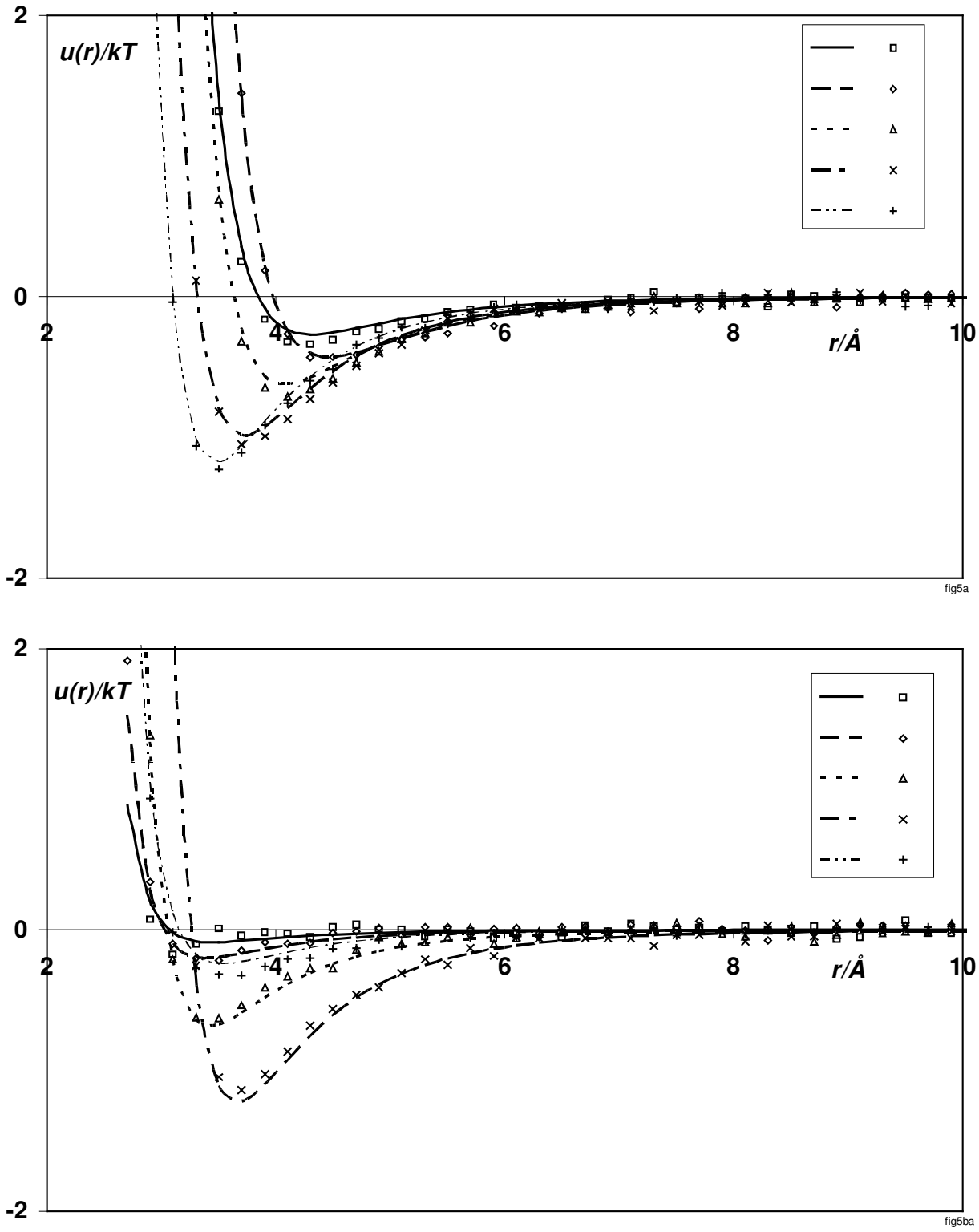


Figure 2.5 Five Lennard-Jones pair potentials reproduced worst from the 90% best established potentials in the training and test sets. The neural network was trained to Lennard-Jones data. a) training set, b) test set, lines: original data, markers: output data

2.4 Conclusions

We proposed a 'quick and dirty' method to derive pair potentials from diffraction data on liquids. The method had no physical background: artificial neural networks were trained on known pair interaction – structure factor pairs of one-component systems. We tested the performance of the networks on different type of potentials, like Lennard-Jones, Morse and Buckingham ones. The training set consisted of 2000 pair potential – structure factor pairs, and the test set was prepared of 200 pairs. We found that the reproduction of the Lennard-Jones potentials and the most of the Morse ones were satisfactory. The test on Lennard-Jones potentials provided quantitatively good reproductions in 95-98%, if the training consisted also Lennard-Jones type interactions. If other type of potentials were included additionally in the training set, the qualitatively good part decreased to 85-90% depending on the extra training sets. Our data were slightly worse for Morse potentials. The performance of the networks was weaker in the case of the Buckingham potentials drastically. It was qualitatively acceptable in 51%, if the training set consisted only of Buckingham pair potential – structure factor pairs. If all type of data were included in the training set, the qualitatively good reproduction of the Buckingham potentials decreased down to 32%.

We hope, an increase of the training set with other type of potentials may lead to a neural network-like generalisation of the inverse problem and it would be an easy way to obtain pair potentials from diffraction data, or at least to obtain reliable starting potentials for iterative potential-refinement procedures. In this study, we focused on systems, where dispersion forces are dominant. For other type of systems (liquid metals, charged colloids…) new trainings of neural networks are necessary, but we think, it is possible to cover the variety of all type of one-component systems with a few trained networks. Therefore, the extension of the training data set and an open server version of the method are in progress by others relying on this work. The open server version of the method determines the pair potential for an experimental structure factor sent on Internet within few seconds on the bases of previously trained networks.

3 Simulation of Binary Hard Sphere Systems with 1:5 and 1:10 Size Ratios

3.1 Introduction

Hard-sphere and hard-body models are often used to model colloid particles in solutions. In the simplest case, the system is treated as a one-component hard-sphere system of colloid particles and the solvent does not appear in the model. This approximation can be used also in the evaluation of different experimental data, e.g. of diffraction experiments (BrunnerPopela and Glatter 1997; Preu et al. 1999; Bergmann et al. 2000; Borbely 2000; Tomsic et al. 2004). The size of the hard spheres can be an individual parameter and this diameter may not have connection to the real diameter of the colloid particles (Pedersen 1994; BrunnerPopela and Glatter 1997; Preu et al. 1999; Bergmann et al. 2000; Borbely 2000; Tomsic et al. 2004). The asymmetric mixture of hard spheres is a more detailed model. Here, the large spheres represent the colloid particles, and the solvent is modelled with the small ones. There are many systems, where the interactions can be satisfactory modelled by these asymmetric binary hard spheres, e.g. colloidal silica particles in different solvents (Imhof and Dhont 1995). The limits of this approximation are studied several times, e.g. in reference (Germain et al. 2003). The task of calculating the structural properties of an asymmetric binary hard-sphere system is extremely large in a computer simulation, if the size ratio of the spheres approaches the colloid/solvent ratio. One can turn the difficulty, if the so-called effective one-component systems are used (Rosenfeld 1994; Dickman et al. 1997; Dijkstra et al. 1999; Clement-Cottuz et al. 2000; Roth et al. 2000; Malherbe and Amokrane 2001). Here an effective (not hard-sphere type) potential is used, where the form of the potential depends on the size ratio ($R=d_l/d_s$) and the volume fraction of the two spheres (η_l and η_s). The volume fraction is defined as $\eta_i=\rho_i\pi d_i^3/6$, where d_i is the diameter and ρ_i is the number density of the given type of spheres. The subscripts s and l denote the small (solvent) and the large (colloid particle) spheres, respectively. The effective potentials are also called as depletion potentials according to the osmotic effect observed in asymmetric binary mixtures (Asakura and Oosawa 1958): when two large spheres approach each other closer than the diameter of the small spheres, the small spheres are excluded from between the large ones. The pressure exerted of the small

spheres on the big ones becomes one sided, and the lack of the pressure between the two large spheres induces an attraction between the two large spheres. One can find several calculations with these effective potentials in the literature (Rosenfeld 1994; Dickman et al. 1997; Dijkstra et al. 1999; Clement-Cottuz et al. 2000; Roth et al. 2000; Malherbe and Amokrane 2001).

The phase behaviour of asymmetric binary mixtures of hard spheres is extensively studied in the literature. There are phase diagrams for different size ratios containing solid/solid, solid/fluid, fluid/solid and fluid/fluid phases for the two components together with possible glassy states. The classical study was the Percus-Yevick (PY) solution for these systems by Lebowitz and Rowlison (Lebowitz and Rowlinson 1964). One can find several attempts by using the binary model or the effective one-component model (Rosenfeld 1994; Dijkstra et al. 1999; Clement-Cottuz et al. 2000). In some of these studies, the performances and the results of the different methods are compared in detail. We would not resume their results or discuss the different equation of states, because our investigation is purely computational, and a comparison with the enormous number of the possible methods points beyond our study. We are aware of the recent progress in this field (Amokrane et al. 2005) and we appreciate the new results. There are also experimental phase diagrams obtained on hard-sphere like colloidal silica particles (Imhof and Dhont 1995). A further question is the possible phase separation (demixing) of the small and the large spheres, where coexistence solid/solid or fluid/fluid phases can be found (Biben and Hansen 1991; Rosenfeld 1994).

There is a special feature of hard-sphere systems. The pressure can be calculated easily, if we know the value of the pair correlation functions at the contact distance (Allen and Tildesley 1987). It results a central role for this quantity in the most of the theoretical methods applied on hard-sphere systems. One can find a large number of equation of states based on different assumptions (Boublik 1970; Mansoori et al. 1971; Malijevsky and Veverka 1999; Santos et al. 1999; Santos et al. 2002; Viduna and Smith 2002). There is also a large variety of methods to calculate the contact values on simple parametric or sometimes on theoretic ways (Boublik 1970; Mansoori et al. 1971; Attard 1989; Henderson et al. 1996; Barrio and Solana 2000; Santos et al. 2002; Viduna and Smith 2002; Viduna and Smith 2002). Contact values were published in many articles and they were compared to each other and to simulation data (Fries and Hansen 1983; Attard 1989; Santos et al. 2002; Viduna and Smith 2002). The theoreticians of simple liquids less focus on the knowledge of the pair correlation functions in the whole r-range, but it is required in

the modelling or evaluation of experimental data, as it was mentioned earlier. There is a bunch of $g(r)$ approximations (Attard 1989; Kahl and Pastore 1991; Yuste et al. 1998) whereof we mention only two: the traditional PY approximation (Lebowitz 1964) and the rational function approximation (RFA) methods (Yuste et al. 1998). We mentioned these two methods from practical cause, since a Mathematica (Wolfram 1999) script (Yuste et al. 1998) can be accessed for their calculation. There are several simulation results too, mostly on the small-small and small-large partial pair correlation functions. One can find studies also on ternary hard-sphere systems (Paschinger et al. 1998; Sindelka and Boublik 1998).

The first simulations on binary mixtures of hard spheres were performed close to 40 years ago. Here, we mention some of the previous computer simulations on highly asymmetric ones without the request of entirety. Fries and Hansen (Fries and Hansen 1983) performed Monte Carlo simulation with 1:2 and 1:3 ratio of radii in 1983. The sums of the packing fractions of the two species were up to 0.6 and the partial pair correlation functions were presented also for the large-large case. We found three set of simulations published in 1997 (Dickman et al. 1997; Malijevsky et al. 1997; Yau et al. 1997). Dickman et al. simulated systems with 1:5 and 1:10 ratios of radii, but the large particles were in infinite dilution. Malijevsky et al. applied smaller asymmetry; the ratio of radii was up to 3.3. The higher ratios of radii were 10 in the study of Yau et al. and the sums of the packing fractions were up to 0.3. (Cao et al. 2000) reaches the 0.5 volume ratio for the small particles, but the systems were dilute for the large particles. They published the simulation on the size ratio of five in 2000. There are also molecular dynamics studies, e.g. a study of the glassy dynamics of binary hard-sphere mixtures up to a size ratio of 1.7 (Foffi et al. 2003). It may be strange, that there are no extensive studies for large packing fractions and asymmetries close to the colloidal limit in the literature, but there are several special computational problems in the simulation of asymmetric binary hard-sphere systems. If one would like to use reasonable packing fractions for both the small and the large particles, it is impossible to choose large asymmetry. Let us demonstrate it in Table 3.1, where we summarised our simulations. The number of the small particles depends cubic on the ratio of the radii. The total number of the particles is also enormous, if η_l/η_s is small. The highest number of particles, where we could get satisfactory results, is about 150 thousand in Table 3.1. In the case of potentials, where the derivative is continuous, this number is treatable especially in Monte Carlo simulations. Unfortunately, Monte Carlo simulations on dense hard-sphere systems like to be

stuck: the acceptance ratio decreases enormously also for very small maximal displacements. In the case of molecular dynamics, there is a substantial difference between hard-sphere and other algorithms. The time step is determined by the next collision for hard-sphere systems, while for continuous potentials it is a predetermined value. For dense systems containing several thousand particles, the collisions are very close to each other in time slowing down the time evolution of the systems. One can find several algorithms to speed up hard-sphere molecular dynamics (Krantz 1993; Marin et al. 1993; Isobe 1999), and the algorithms are also compared to each other (Krantz 1993). We realised during our study, as yet there is not any method to reduce the computation time with magnitudes.

We started our study on the initiation of a colloid chemist colleague who would like to get large-large pair correlation functions of binary hard-sphere systems at different packing fractions calculated by simulations. We checked the literature and we found that there is a lack of simulation data, at least for dense systems. This was the main initiation to start our calculation and it seemed to be possible due the recent development of the computers. Besides this demand on the pair correlation functions, we were interested in other questions, too. We would like to quantify the inappropriateness of the one-component hard sphere model in colloid science, especially because of its role in the evaluation of experimental data. We obtained partial pair correlation functions in our simulations therefore we could compare them to theoretical ones. We were especially interested in the power of some selected theoretical methods for the determination of large-large pair correlation functions. At last, our simulations provided a new set of contact values.

3.2 Details of the Calculations

According to the simplicity of the Monte Carlo methods opposite to molecular dynamics, we started with Monte Carlo methods. Unfortunately, the calculations were very inefficient. The acceptance ratios decreased drastically in the dense systems. We were stuck in most cases: small acceptance ratios were possible already at very small maximal moves. We changed to molecular dynamics, because collective motion seemed to be necessary.

In the case of dynamics, the algorithm contains few important steps: I) search for the next collision time, II) fly of all particles till this time, III) perform the collision and IV) calculate the pair correlation functions. In a non-optimised program, steps I and IV scale with the number of

particles quadratic, while steps II and III scale linearly. One can find many attempts in the literature to reduce the quadratic scaling. Unfortunately, all attempts seemed to be efficient only for a given size range of systems at given density ranges and for the specific features of the applied computers (e.g. memory/CPU speed, vector possibilities). Finally, we elaborated our own algorithm optimised for the planned system sizes and the available computers.

We used a link cell method to search the next collisions between small-small and large-small particles. The cell size was determined as the sum of small and large sphere radii multiplied with a factor. The factor was optimised to minimise the computation time, and it was around 1.11-1.22. The large-large collision times were calculated without cell lists. An optimised number (850) of next collisions were stored. The stored collision times were updated only for the particles taken part in the last collision. The total search of the possible collisions was performed, only if the maximum displacement of the particles might imply incompleteness in the collision list or there were no collision times from the last total collision search. These total calculations were performed after 21-614 collisions in average. The list of which particle belonged to which cell was recalculated after each collision.

The time consuming calculation of the pair correlation functions were performed only after each 200000 collisions, except for the large-large case, where it was calculated after each collision.

Our original idea was to exploit the systems with 1:10 and 1:20 size ratios, but finally we had to reduce it to the 1:5 and 1:10 size ratios, because the calculation of 1:20 ratio exceeded our computational possibility. The packing fraction of the large particles was between 0.1-0.5 and 0.1-0.4 was used for the small particles. We put 100 large particles in the simulation cells, mostly. The number of the small particles was between 2500 and 150000, where we got reasonable results.

In the cases, where η_l was small and η_s was large, the number of the particles was too large. Here, we reduced the number of the large particles to 50. Finally, where we couldn't produce reasonable results with 50 large particles in reasonable time, we reduced the number of the large particles down to 25. It did not help us produce better results, maybe 25 large particles means too weak statistics. There are two further effects hindering the simulation of dense systems. The increased density shortens the time till the next collision. Furthermore, small η_l/η_s increased the number of the particles, and the time till the next collision is reciprocal to the number of particles

anyway. Therefore, we could not perform reliable simulations in some cases. The simulations are summarised in Table 3.1.

	$\eta_l\downarrow$ $\eta_s\rightarrow$	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
1:5 diameter ratio	**0.1**	100/0	100/12,500	100/25,000	100/37,500 100/37,500 50/18,750	100/50,000 50/25,000	100/62,500 *100/62,500* 50/31,250 50/31,250
	0.2	100/0	100/6,250	100/12,500	100/18,750	100/25,000	
	0.3	100/0	100/4,167	100/8,333	100/12,500		
	0.4	100/0	100/3,125	100/6,250 100/6,250			
	0.5	100/0	100/2,500 100/2,500				
	$\eta_l\downarrow$ $\eta_s\rightarrow$	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
1:10 diameter ratio	**0.1**	100/0	*100/100,000* 50/50,000 50/50,000 *25/25,000*	*50/100,000* *25/50,000*	*50/150,000* *25/75,000*	*100/400,000* 25/100,000	*20/100,000* *25/125,000* *25/125,000*
	0.2	100/0	100/50,000, 50/25,000 50/25,000	50/50,000 50/50,000	*50/75,000* 50/75,000	*50/100,000* *50/100,000*	*50/125,000*
	0.3	100/0	100/33,333 50/16,666	100/66,666 50/33,333 50/33,333	100/100,000 50/50,000 50/50,000 50/50,000	*100/133,333* 50/66,666	
	0.4	100/0	100/25000 100/25000	100/50000 50/25000	200/150,000 100/75000 50/37500 50/37500		
	0.5	100/0	100/20000 100/20000	100/40000 100/40000			

Table 3.1 System sizes of the performed simulations. The systems were printed in italics, if we were not satisfied with the achieved statistical confidence of the pair correlation functions according to the limited computation time or possible glassy phases.

The pair correlation functions were calculated with a grid of 0.05, where the diameter of the small particles was set to one.

Many theoretically important properties were not calculated in our simulations, because we focused on our purposes detailed at the end of the introduction. For example, we did not check the phases of our systems or the self-diffusion coefficients, because they could be done only by a drastic increase of the computational time. In the cases, where the pair correlation functions showed solid or glassy features, we started control simulations from different starting configurations (Table 3.1). If the data were close to each other, we averaged them. The calculated pair correlation functions were averaged over 10-100 million of collisions after the equilibration.

3.3 Results and Discussions

3.3.1 Quantitative Limits of the One-Component Hard-Sphere Model

As we detailed in the introduction, the one-component hard-sphere model is often used in the modelling of colloid systems and in the computer-aided evaluation of experimental data on colloids. In spite of the widespread use of it and everybody's suspicion on its failure, there is a lack of simulation study up to now. There is a theoretical study of Heno and Regnaut (Heno and Regnaut 1991), where they showed that the PY solution of binary hard spheres at the colloid limit differs from the PY solution of one-component hard-spheres. Our simulations quantitatively tested the one-component approximation. Two cases are shown in Figure 3.1 and Figure 3.2. The large-large pair correlation functions are presented for systems at η_l=0.3 and η_s=0-0.3 at the 1:5 size ratio in Figure 3.1. The presence of the 'solvent' changed the feature of the pair correlation functions drastically. The systems containing small particles were ordered more, further peaks appeared in the $g_{ll}(r)$-s and the height of the first peaks became extremely higher. The amount of the solvent was important not only for the height of the peaks but also for their r-position. Higher amount of solvent brought closer the coordination spheres to the central particle. In Figure 3.2 we present data where the total packing fraction was kept fixed: $\eta_l+\eta_s$=0.5. The one-component system differed reasonably from all the two-component ones, while the two-component $g_{ll}(r)$-s were relatively close to each other. In Figure 3.1 we showed the situation, when one simple forgot the solvent and used the real particle radius of colloids in the one-component hard-sphere approximation. In Figure 3.2 the situation was shown, where one used a different density for the

inter-particle structure modelling than the estimated one for the colloid particle. The figures emphasise graphically that both approximations are rather misleading.

We would like to quantify these approximations. An overlap integral was defined between two pair correlation functions, $g_1(r)$ and $g_2(r)$.

$$O = \frac{\sum_{i=1}^{b} g_1(r_i) g_2(r_i)}{\sqrt{\left(\sum_{i=1}^{b} g_1(r_i)^2\right)\left(\sum_{i=1}^{b} g_2(r_i)^2\right)}} \qquad \textbf{(3.1)}$$

where b was the number of the bins in the r-space. O was between zero and one, a value close to one meant similar $g(r)$-s. The overlap integrals are shown in Table 3.2. Two cases are presented for each size ratio. In the first case, the reference pair correlation functions reflected the situation in Figure 1, where the binary systems were compared to the one-component ones at the same packing fraction of the large particles. In the second case, the binary systems were compared to that one-component system, where the one-component packing fraction equalled to the sum of the binary packing fractions. The similarity of the $g(r)$-s behind the numbers could be estimated, if the corresponding data in Figure 3.1 and Figure 3.2 and in Table 3.2 were searched. The closest binary $g(r)$ to the one component system had an O=0.923 in Figure 3.1 and O=0.922 in Figure 3.2. The trends were clear in Table 3.2. In the first case, where η_l(binary)=η_l(one-component), the presence of increasing amount of 'solvent' destroyed the similarity, especially at increasing packing fraction of the large particles. In the second case, where η_l(binary)+η_s(binary)=η_l(one-component), all of the binary systems were far from the one-component one. Here, the overlap integral decreased, if the solute was replaced by solvent.

There were differences also in the data sets of the 1:5 and 1:10 simulations. Of course, we should mention in accordance with Table 3.1, that the statistical confidence of the 1:10 simulations was reasonable weaker than those of the 1:5. It meant, the overlap integrals were waited to be slightly smaller than for the 1:5 ratio. In spite of this, the differences between the two sets implied that the one-component approximation is cruder for the systems towards the colloid limit. It means, for real size ratios of a few Å large solvent / nanometer large colloid particles, the use of this approximation is not suggested in accordance with the theoretical findings (Heno and Regnaut 1991).

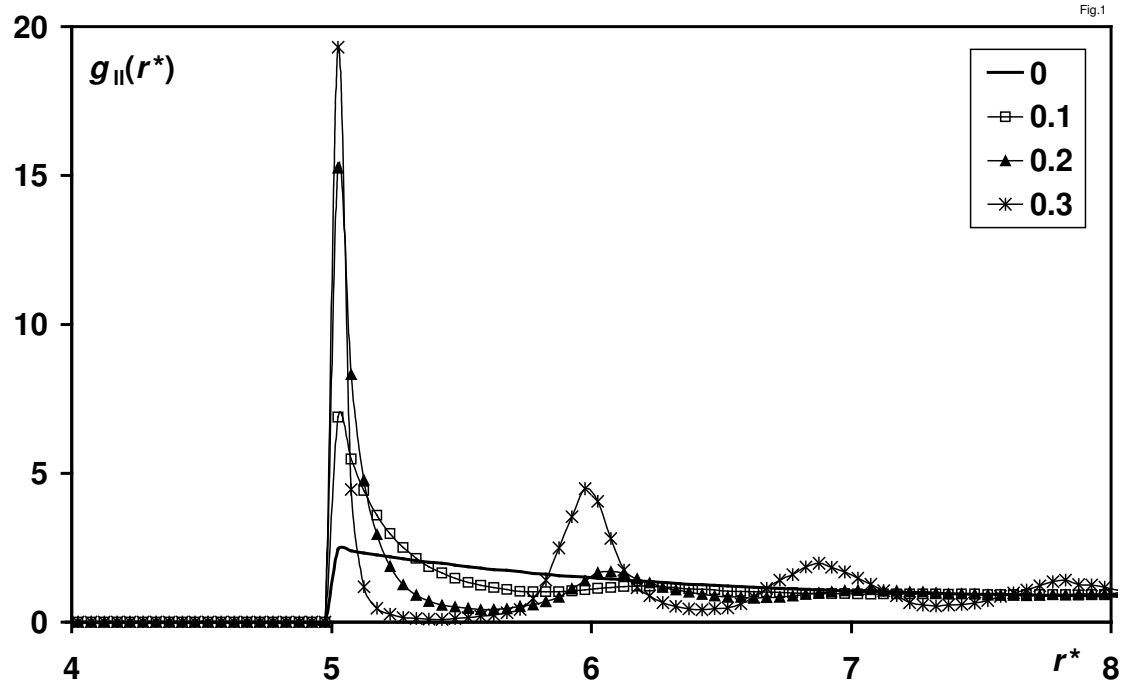


Figure 3.1 Comparison of large-large partial pair correlation functions for the systems of η_l=0.3 and η_s=0-0.3 at size ratio of 1:5. The legend shows η_s.

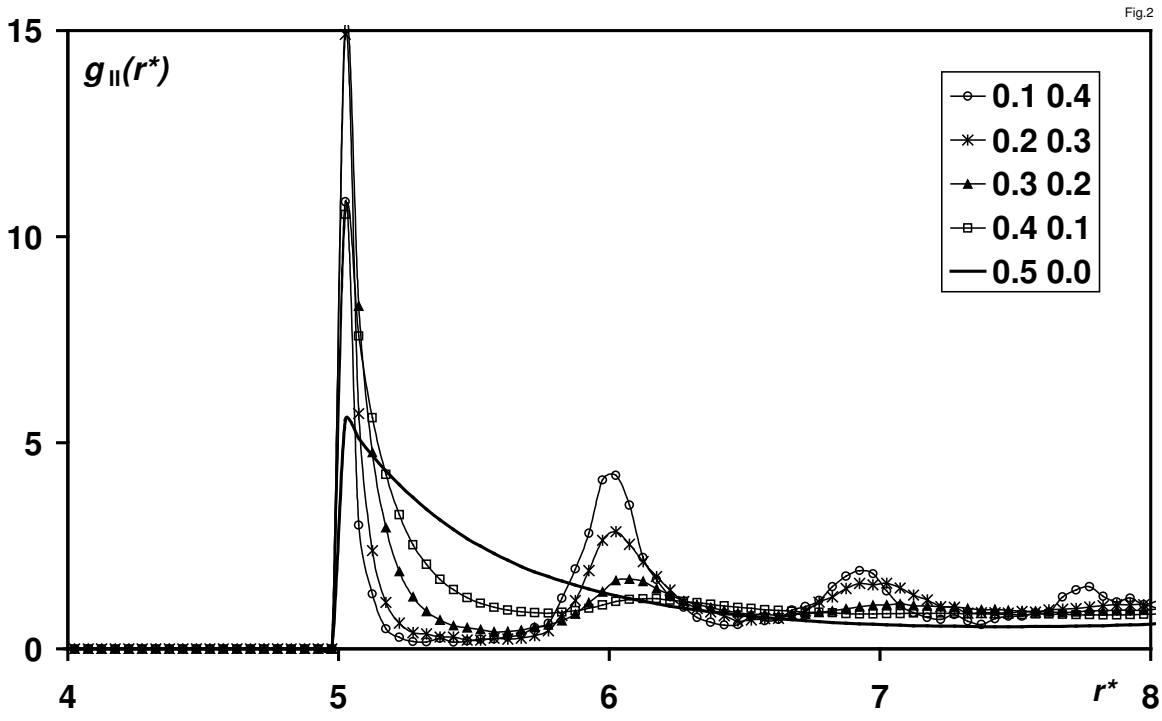


Figure 3.2 Comparison of large-large partial pair correlation functions for the systems of η_l+η_s=0.5 at size ratio of 1:5. The first and second columns in the legend show η_l and η_s, respectively.

The most important phenomenological cause of our findings is the existence of the depletion force. We did not show any data on the small-small and small-large partials, because they were not involved in the question of this subsection. Anyway, these partials were less sensitive to the change of the packing fraction parameters.

	1:5 diameter ratio					1:10 diameter ratio			
Overlap integrals if, $\eta_l=\eta$ of the one-component reference									
$\eta_l\downarrow$ $\eta_s\rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.1**	**0.2**	**0.3**	**0.4**
0.1	0.97	0.85	0.73	0.75	0.72	**0.96**	*0.89*	*0.86*	*0.73*
0.2	0.96	0.80	0.70	*0.66*		0.86	0.84	*0.84*	*0.91*
0.3	0.92	0.72	0.62			0.84	0.73	0.79	
0.4	0.88	0.65				0.74	0.61		
0.5	0.83					0.65			
Overlap integrals if, $\eta_l+\eta_s=\eta$ of the one-component reference									
$\eta_l\downarrow$ $\eta_s\rightarrow$	**0.1**	**0.2**	**0.3**	**0.4**		**0.1**	**0.2**	**0.3**	**0.4**
0.1	0.98	0.88	0.79	0.72		0.97	*0.88*	*0.81*	*0.58*
0.2	0.97	0.86	0.74			0.89	0.85	*0.77*	
0.3	0.95	0.80				0.86	0.76		
0.4	0.92					0.79			

Table 3.2 Overlap integrals to quantify the limit of the one-component hard-sphere approximation for two cases: a) $\eta_l=\eta$ of the one-component reference b) $\eta_l+\eta_s=\eta$ of the one-component reference

3.3.2 Comparison to Percus Yevick and Rational Function Approximation Pair Correlation Functions

There are many theoretical approximations to calculate the pair correlation functions of binary hard-sphere systems. We chose two of them to compare to our simulation data in order to quantify their ability. The first was the PY method on binary hard-spheres (Santos et al. 1999) that importance is not only historical. The other one was the RFA method developed by Santos et al. (Santos et al. 1999). We chose these two methods from practical cause, since we could use the Mathematica files provided by these authors (Yuste et al. 1998). We know that there is a bunch of well-developed methods [see references in (Amokrane et al. 2005)]. We just used these data where $r>d$, below that we assumed the $g(r)$-s to be zero.

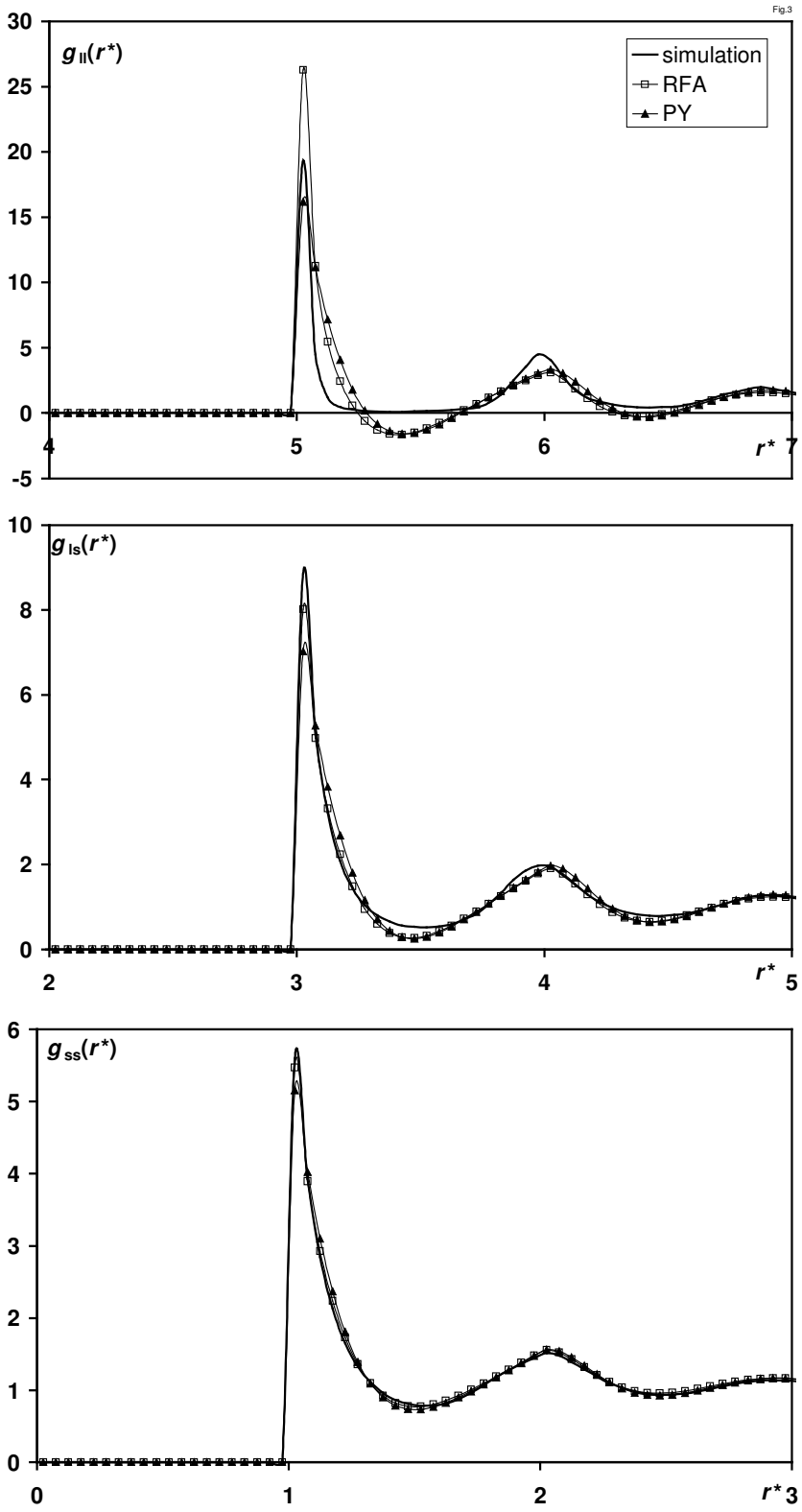


Figure 3.3 Comparison of the pair correlation functions obtained with simulation and theoretical methods for the system of η_l=0.3 and η_s=0.3 at size ratio of 1:5.

The theoretical methods performed well for the small-large partials and especially for the small-small partials (Figure 3.3). Unfortunately, the case of the large-large partial was not so satisfactory. For example, one can find unphysical minima (negative $g(r)$) after the first peaks at high packing fractions. We mention here, that there are theoretical approximations where these unphysical parts cannot be detected, but the choice of the PY and the RFA methods was practical. The dependence of the large-large partials on the packing fraction of the 'solvent' is presented in Figure 3.4. If η_l=0.3, the theoretical RFA $g(r)$ was satisfactory for η_s=0.1. The failure of the theory increased up to drastically misleading and unphysical functions above this density. We presented the data only for the RFA method in Figure 3.4, but those of the PY were very similar.

To quantify the confidence of the theoretical approximations, we calculated the overlap integrals between the theoretical and simulated curves of the systems at same packing fractions. Let us consider the 1:5 ratio first. The overlap integrals were very close to one (0.997-1.000 for RFA and 0.996-1.000 for PY) for the small-small partials. In the case of the small-large $g(r)$-s, the agreement was satisfactory (0.986-1.000 for RFA and 0.976-1.000 for PY). We found small trend in respect the size ratio and the packing fractions. Higher packing fractions of the large particles and especially of the small particles implied some deviations. The situation was not good for the large-large partials as it can be seen in Table 3.3 for the simulation-RFA overlap integrals. Higher packing fraction of the large particles destroyed the similarity of the curves. It was more pronounced, if the packing fraction of the solvent increased, too. The differences originated mostly from the first part (first few maxima and minima) of the theoretical $g(r)$-s. The RFA data were between 0.867-1.000 and the PY ones were between 0.838-1.000. To estimate the link between the visual and numerical comparisons, O_{ll}=0.949 for RFA and O_{ll}=0.897 for PY in Figure 3.3.

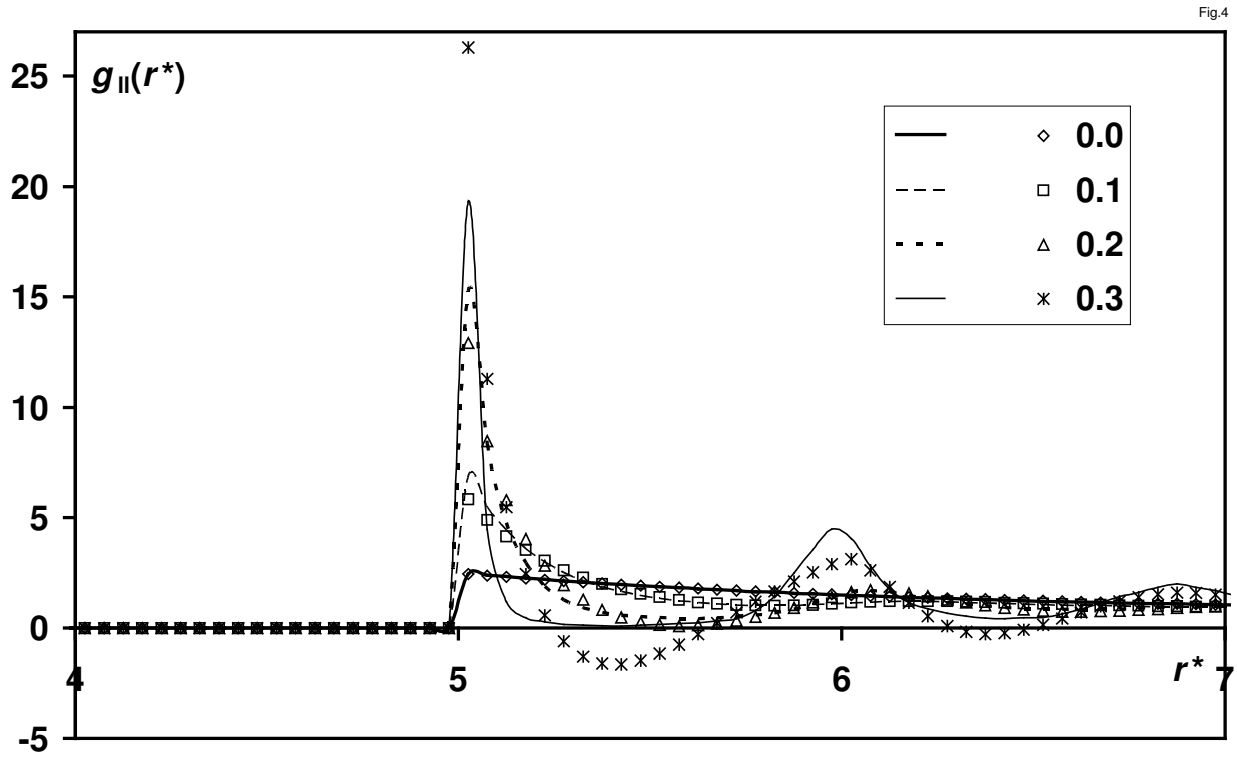


Figure 3.4 Comparison of large-large partial pair correlation functions obtained by simulations and the RFA method for the systems of η_l=0.3 and η_s=0-0.3 at size ratio of 1:5. The legend shows η_s.

We did not present the comparison between the PY method and the simulations and between the two theoretical methods in detail. The difference between the PY and the RFA sets was small, except the large-large partials (RFA-PY overlap integrals: O_{ss}=0.999-1.000, O_{sl}=0.996-1.000, O_{ll}=0.957-1.000). It is clear from Table 3.3, that in comparison to the simulations the PY method performed weaker than the RFA in average, and it might be termed as a significant difference between the two approximations, especially at high densities.

As it was detailed, the statistical confidence in our 1:10 data was weaker than of the 1:5 data. The overlap integrals were between 0.995-1.000 (RFA) and 0.994-1.000 (PY) in the small-small partials. They were between 0.986-1.000 (RFA) and 0.984-1.000 (PY) in the small-large partials. The large-large data are shown (0.714-1.000) for RFA in Table 3.3, while it was 0.805-1.000 for PY. We found that in some cases at the 1:10 diameter ratio PY performed better than RFA, but maybe it can be attributed to the weaker statistics of the 1:10 data. In spite of taking into account the bias of the weak statistics, we think the theoretical curves showed less similarity to the

simulation data in the 1:10 case than in the 1:5 one. It means a decrease in the efficiency of these theoretical methods towards the colloid limit. It coincides with the theoretical trend of the first minimum of the $g_{ll}(r)$ in the PY approximation; it goes to minus infinity in the colloid limit.

A comparison of Table 3.1 and Table 3.2 shows that the one-component hard-sphere approximation is expressively weaker than the use of theoretical binary hard-sphere curves. Approximate theories performed better than a simple but bad choice of reference system. Of course, the unphysical parts of the PY and RFA $g(r)$-s in the latter should warn to use them cautiously.

	1:5 diameter ratio						1:10 diameter ratio				
$\eta_l\downarrow$ $\eta_s\rightarrow$	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0**	**0.1**	**0.2**	**0.3**	**0.4**
0.1	1.0000	0.9995	0.9918	0.9877	0.8976	0.8675	1.0000	0.9872	*0.7737*	*0.7561*	*0.1845*
0.2	1.0000	0.9990	0.9940	0.9701	*0.9103*		1.0000	0.9904	0.9093	*0.7143*	*0.3339*
0.3	1.0000	0.9976	0.9899	0.9487			1.0000	0.9886	0.9105	0.7725	
0.4	1.0000	0.9965	0.9824				0.9999	0.9820	0.8595		
0.5	0.9996	0.9939					0.9996	0.9707			

Table 3.3 Overlap integrals to quantify the theoretical methods. *O* is calculated between the simulated and RFA pair correlation functions of the same systems.

3.3.3 Contact Values

The contact values, $g(d)$-s, of the pair correlation functions play central role in hard-sphere systems. The pressure of these systems can be calculated easily, if one knows the contact data. In the one component case, $\frac{P}{\rho kT} = 1 + \frac{2\pi\rho d^3}{3} g(d)$, where ρ is the number density, d is the diameter of the spheres, T is the temperature and k is the Boltzmann-constant. In the case of MC simulations, the contact value is determined usually by extrapolation to the contact distance (Allen and Tildesley 1987). This extrapolation can be performed on $g(r)$ or on $\ln g(r)$ (Sindelka and Boublik 1998). In the case of molecular dynamics of hard-spheres, an opposite direction is possible, too. One can calculate the pressure in the knowledge of the positions and impulses during the collisions. In spite of using molecular dynamics, we calculated the contact values with extrapolation of $\ln g(r)$. We chose this way to avoid any extra calculations during the simulations

and the pair correlation functions were calculated according to the other purposes of our study, anyway. We mention here a possible failure of pair correlation functions calculated by hard-sphere molecular dynamics. The snapshots of the molecular dynamics calculations are taken usually at the time of collisions. Therefore, the contact distance is over-represented in the pair correlation functions. Depending on the resolution of the $g(r)$ it can be visually detected for systems consisting up to a few hundreds of particles. The effect is negligible for large systems. In the case of small systems, a correct solution is for example to take the snapshots not at the collisions but at equidistant times during the simulation.

Table 3.4 contains our extrapolated contact values. The estimated error of the values was between 0.2-5% for the small-small and small-large partials, where the upper limit corresponded to the high packing fraction of the solvent particles. At the diameter ratio 1:10 the error was even larger for the systems printed in italics (Table 3.4). The estimated error was up to 25% for the contact value of the large-large partials, and the error was about 50% or even more for the data printed italics in Table 3.4.

The contact values of the small-small and small-large partials are in good agreement to the data can be obtained by the PY or RFA approximations. In the case of the large-large partials, our data are generally larger. This effect was more pronounced for large packing fractions.

Finally we discuss some visual observations on the large-large $g(r)$-s. As we mentioned earlier, in some of our systems the solute, the solvent or both were in solid state. In these cases the $g(r)$-s showed many and usually not regular sharp peaks. We found a special type of characteristic peaks in the $g_{ll}(r)$-s also at densities corresponding to liquid phase. These peaks reflected the situation where layers of solvent particles were sandwiched in then space between of two solute particles. Entropic effects caused these layers, if the distance of two solute particles was larger than the diameter of the solvent particle. Below this distance the depletion effect forced the solute particles to contact each other. At the diameter ratio 1:10 one solvent layer was observed at η_l/η_s=0.2/0.2; 0.3/0.2; 0.5/0.1, two layers were observed at η_l/η_s=0.1/0.3; 0.2/0.3; 0.2/0.4; 0.5/0.2 three layers were observed at η_l/η_s=0.3/0.3; 0.4/0.2. In the other cases, the peaks ran into one another. The peaks of these layers were less pronounced for the 1:5 diameter ratio, thus the surfaces between the solvent molecules can be sandwiched are smaller than at the 1:10 diameter ratio. An example of the sandwiched particles is shown in Figure 3.5. Two simulations are presented, where a pronounced layering can be seen on the data of η_l=0.1 and η_s=0.4 simulation.

On the basis of simple idea of increasing density means stronger structural order, one should guess a more ordered structure for the η_l=0.4 and η_s=0.1 densities in the large-large partial.

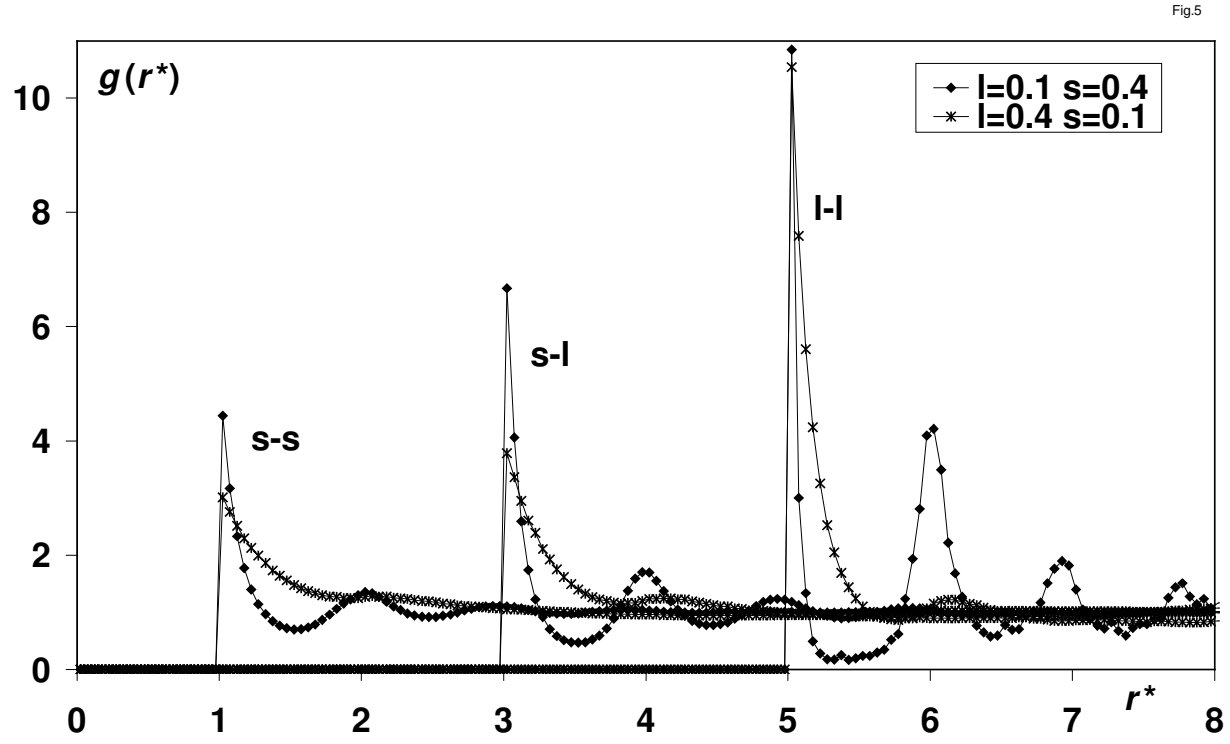


Figure 3.5 Partial pair correlation functions for two simulations at size ratio of 1:5.

We performed a rather large number of simulations thereof the data might be interesting for theoreticians to compare the different approximations and for experimental researchers to use them in data evaluation of diffraction data.

	1:5 diameter ratio						1:10 diameter ratio				
					large-large						
$\eta_l\downarrow$ $\eta_s\rightarrow$	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0**	**0.1**	**0.2**	**0.3**	
0.1	1.30	3.32	9.21	18.52	20.60	*40.70*	1.31	4.45	*4.25*	*26.54*	
0.2	1.75	4.94	12.53	24.08	*46.02*		1.73	10.25	14.06	*16.23*	
0.3	2.47	7.72	20.66	40.17			2.46	12.08	25.62	7.74	
0.4	3.65	12.43	34.34				3.71	23.24	49.94		
0.5	5.80	21.76					6.35	46.33			
					large-small						
$\eta_l\downarrow$ $\eta_s\rightarrow$		**0.1**	**0.2**	**0.3**	**0.4**	**0.5**		**0.1**	**0.2**	**0.3**	**0.4**
0.1		1.77	2.81	4.65	8.55	*14.96*		1.79	*2.61*	*4.83*	*13.84*
0.2		2.20	3.76	6.96	*14.40*			2.17	3.73	*6.96*	*2.36*
0.3		2.92	5.37	11.57				2.77	5.18	9.66	
0.4		4.01	8.70					3.81	8.38		
0.5		6.31						5.57			
					small-small						
$\eta_l\downarrow$ $\eta_s\rightarrow$		**0.1**	**0.2**	**0.3**	**0.4**	**0.5**		**0.1**	**0.2**	**0.3**	**0.4**
0.1		1.54	2.18	3.22	5.26	*9.00*		1.52	*1.95*	*3.20*	*88.36*
0.2		1.87	2.80	4.44	*8.03*			1.81	2.67	*4.28*	*10.12*
0.3		2.41	3.79	6.69				2.22	3.45	5.92	
0.4		3.15	5.57					2.91	5.07		
0.5		4.52						3.99			

Table 3.4 Contact values of the pair correlation functions obtained by the simulation.

3.4 Conclusions

The main goal of this study was to get a whole set of pair correlation functions of binary hard-sphere mixtures at 1:5 and 1:10 size ratios for a wide and systematic range of packing fractions. This calculation was initialised by the demand on these functions in the modelling of colloid science. Besides this data accumulation task we were interested in some features connecting to the $g(r)$-s obtained.

Our first question concerned the limit of the approximation of two-component hard-sphere structure with one-component hard-sphere one. Besides the usual qualitative visualisation of the pair correlation functions, we calculated overlap integrals between various pairs of $g(r)$-s to

quantify this limit. We checked two approaches. In the first case, binary systems were compared to one-component systems, where the packing fractions of the large particles were the same in both systems. We found, that this approximation cannot be applied, if $0.2 \leq \eta_l$. The increase of η_s decreases the similarity of the $g_{ll}(r)$-s, too. One and two-component pair correlation functions were compared also in a second approach, where the sum of the packing fractions in the binary systems was equal to the one-component η. This approach is questionable also at small η_s, the presence of any 'solvent' reduced the similarity of the $g(r)$-s drastically. On the other hand, the $g_{ll}(r)$-s of the binary systems with an equal sum of packing fractions are rather close to each other. It implies a possibility of a loose choice of system parameters in the modelling of colloids or in the computer-aided evaluation of experimental data on colloids.

We studied the performance of the PY and the RFA theoretical methods in the creation of pair correlation functions. These two methods were chosen only from a practical viewpoint. They could be easily calculated with computer script provided by the authors. The theoretical $g(r)$-s were compared to the simulation ones both graphically and quantitatively. The performances of these theoretical methods were convincing in the small-small and small-large cases. On contrary, the large-large pair correlation functions lose their physical reality at medium and large densities. The two methods behave rather similarly, but the PY method seemed to be worse than the RFA method. The large-large partials of these analytical functions failed for dense ($0.2 < \eta_s$ or $0.5 < \eta_s + \eta_l$,) systems.

In the comparison of the 1:5 and 1:10 data, we concluded that both the approximation of the one-component hard-sphere model and the theoretical models are less successful as we approach the colloid limit.

Finally, we reported a set of contact values of $g(r)$-s for the binary hard-sphere systems studied by molecular dynamic simulations.

4 Parameterization of Coulomb Interaction in Three-Dimensional Periodic Systems

4.1 Introduction

The calculation of long-range interactions, e.g. Coulomb interactions, in periodic systems is a long-standing question in different areas of physics. The progress in computational modelling reassigned it partly to a numerical mathematical task, but the different physical ideas still play important roles.

For the energy calculation of a long-range potential the trivial method is to sum the interactions among all particles in all cells of the systems. If N denotes the number of the cells taken into account, and n is the number of the particles within one cell, the trivial energy calculation of the total system scales with N^2n^2. If one uses the symmetry of the periodic images, and one is interested in the potential energy of a central cell, as it is schematically shown in Figure 4.1, the trivial calculation scales with Nn^2. This calculation is unfeasible, because one should apply a large number of image cells. A recommended minimal choice of N is a few thousand. The first computationally essential milestone in making it feasible was the method of Ewald (Ewald 1921). He suggested recomposing the interactions into two main parts. The direct sum is given by the sum of screened interactions, using Gaussian-screening functions centred at each point charge. The reciprocal sum is given by the Fourier series representation of the solution to Poisson's equation in periodic boundary conditions, using the sum of the compensating Gaussians, again centred at each point charge, as a source distribution. The terms can be expressed analytically in these Fourier series (Ewald 1921; Sagui and Darden 1999).

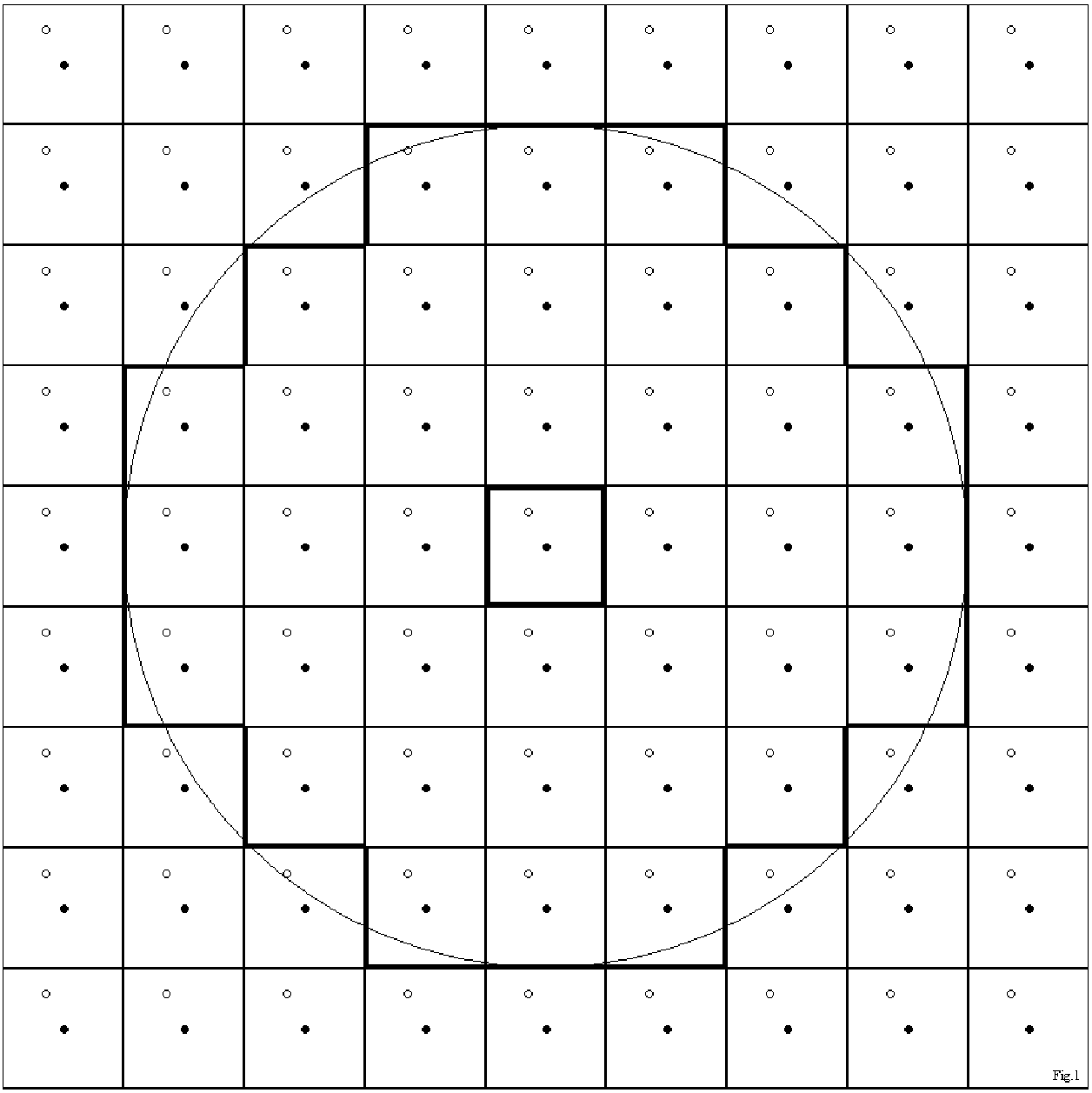


Figure 4.1 Schematic representation of a periodic system. The central cell and a possible choice of an *N* cell subsystem are denoted by thick borders.

The spreading of computer simulations demanded many new approaches especially on the field of liquid matter science. Without going into details and classifying the methods, the most important ones are the reaction-field method (Friedman 1975; Neumann and Steinhauser 1980), the different mesh methods (Eastwood et al. 1980; Darden et al. 1993; Essmann et al. 1995) and

the multi-pole expansions (Ladd 1977; Ladd 1978; Greengard and Rokhlin. 1987). The description of these techniques can be found in the original publications, later applications and in textbooks (Allen and Tildesley 1987; Rapaport 1995; Frenkel and Smit 1996) as well. There are numerous comparative studies, whereof we give references only to some traditional articles (Esselink 1995; Luty et al. 1995; Pollock and Glosli 1996). The methods scale with n differently, but in practical applications not the scaling, but the overall computational cost and accuracy are the most important factors at a given system size. For example, there is an n scaling method (Greengard and Rokhlin. 1987), but the overall computational cost becomes beneficial only for large systems.

One can find also methods derived only using practical and computational assumptions instead of physical background. The simplest is the decomposition of the interaction into pair-interactions and tabulating the potential function in a three-dimensional grid. It is possible also to parameterise the tabulated grid using appropriate functions. This technique was applied in the historical work of Brush et al. (Brush et al. 1966) on one-component plasma. Hansen (Hansen 1973) used cubic harmonic functions (see also kubic harmonics) in the tabulation of interactions. The coefficients of this type of series expansion and the efficient choice of the function series were revisited and discussed several times (Slattery et al. 1980; Adams and Dubey 1987; Anderson et al. 2000). The main advantage of the cubic harmonic functions is the natural possibility of describing cubic symmetry. The coefficients and the corresponding functions of Adams and Dubey (Adams and Dubey 1987) seem to be a practically feasible way in the calculation of Coulomb interactions.

The choice of the method depends on many factors. Practically, the system size is not the most important factor. In the case of classical mechanical molecule models, it is not necessary to assign charges to all species. In this case, the computational cost of the short-range interactions can become dominant against that one of the long-range interactions. The organisation of the short-range interaction calculation differs from the long-range one. They are calculated on interaction pairs via one by one, while the Coulomb interactions are calculated for the whole system simultaneously in the Ewald method. Therefore, if the number of particles involved only in short-range interactions is significant, the program structure should be optimised for the short-range interactions. Another factor is the simplicity of the code. For a program code developed for a few applications in a specific research project, it is not necessary to optimise ultimately for the

speed performance, it is more important to write the code quickly and to be able to eliminate programming failures.

The purpose of our study was to develop a very simple method for the calculation of Coulomb interactions in three-dimensional periodic systems. Simple means that it could be embedded easily into codes written for short-range interactions and it could be at least as simple and efficient than the use of cubic harmonics. Of course, we had to suggest a method that is feasible in the side of calculation cost, as well.

4.2 Partitioning of the Potential Energy in Periodic Systems

The primary aim of our investigation was to parameterise the periodic Coulomb interactions. Before we started the calculations, we checked how the energy of one cell in a periodic system should be calculated. This is not only a case of an arbitrary definition, because the calculated potential energies are used very often. One example is classical mechanical Monte Carlo simulation, where the energies appear in the probability factor. An incorrect energy definition doubles or halves the apparent temperature of the system. The question of the energy partition is also important in micro-canonical molecular dynamic simulations on periodic systems. We discuss this question in this section.

Let us consider a periodic system. The particles interact with a pair-wise additive potential, where the potential depends solely on the distance of the particles. The simulation cell contains n particles. We define a system of all together N-1 periodic images of the central cell and the central one. The schematic visualisation of an N cell system can be seen in Figure 4.1. The image cells are included in the system, if the centre of an image cell is within a given distance from the central cell (Figure 4.1). The potential energy of the total N cell system (U^{tot}), e.g. the electrostatic energy defined by the Coulomb interaction, can be decomposed into three terms. The terms are: the interactions within the cells (intra term), the interactions among the images of the same particles in the different cells (self term) and the interactions of the distinct particles in the different cells (distinct term).

$$U^{\mathrm{tot}} = \frac{1}{2}\sum_{k=1}^{N}\sum_{i=1}^{n}\sum_{j\neq i}^{n} u_{ij}^{kk} + \frac{1}{2}\sum_{k=1}^{N}\sum_{l\neq k}^{N}\sum_{i=1}^{n} u_{ii}^{kl} + \frac{1}{2}\sum_{k=1}^{N}\sum_{l\neq k}^{N}\sum_{i=1}^{n}\sum_{j\neq i}^{n} u_{ij}^{kl} \qquad \textbf{(4.1)}$$

where u_{ij}^{kl} denotes the interaction between the i-th particle in the k-th cell and the j-th particle in the l-th cell. One can count the intra-, self- and distinct interactions:

$$\frac{1}{2}Nn(n\text{-}1) \qquad \frac{1}{2}N(N\text{-}1)n \qquad \frac{1}{2}N(N\text{-}1)n(n\text{-}1)$$

If one collects the interactions where particle one and all of its images are involved, one gets

$$U_1^{\text{all}} = \sum_{k=1}^{N}\sum_{j\neq 1}^{n} u_{1j}^{kk} + \frac{1}{2}\sum_{k}^{N}\sum_{l\neq k}^{N} u_{11}^{kl} + \sum_{k}^{N}\sum_{l\neq k}^{N}\sum_{j\neq 1}^{n} u_{1j}^{kl} \tag{4.2}$$

and

$$N(n\text{-}1) \qquad \frac{1}{2}N(N\text{-}1) \qquad N(N\text{-}1)(n\text{-}1)$$

It is assumed that $u_{1j}^{kl} = u_{j1}^{kl}$ in the last term. If the collection is restricted to the interaction of the first particle in the first cell:

$$U_1^1 = \sum_{j\neq 1}^{n} u_{1j}^{11} + \sum_{l\neq 1}^{N} u_{11}^{1l} + \sum_{l\neq 1}^{N}\sum_{j\neq 1}^{n} u_{1j}^{1l} \tag{4.3}$$

and

$$(n\text{-}1) \qquad (N\text{-}1) \qquad (N\text{-}1)(n\text{-}1)$$

The last equation corresponds to all interactions involved the first cell explicitly:

$$U^1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} u_{ij}^{11} + \sum_{l\neq 1}^{N}\sum_{i=1}^{n} u_{ii}^{1l} + \sum_{l\neq 1}^{N}\sum_{i=1}^{n}\sum_{j\neq i}^{n} u_{ij}^{1l} \tag{4.4}$$

and

$$\frac{1}{2}n(n\text{-}1) \qquad (N\text{-}1)n \qquad (N\text{-}1)n(n\text{-}1)$$

The energy of one cell, e.g. in the Ewald sum, is defined as U^{tot}/N in most of the textbooks (Allen and Tildesley 1987; Rapaport 1995; Frenkel and Smit 1996). On contrary, in most of the lectures at conferences, if one is interested solely on the central cell, the energy of Equation (4.4) is frequently used. Another introduction of the potential energy of periodic systems uses the formulae of electric fields. The effect of charges in the central cell and in the surrounding ones is expressed in an electrostatic potential. The equation is simply

$$U^1 = \frac{1}{2}\sum_{i=1}^{n} q_i \varphi(\mathbf{r}_i) \tag{4.5}$$

where q_i is the charge of the particle and $\varphi(\mathbf{r}_i)$ is the electrostatic potential at the $\mathbf{r}_i$ point. Equation (4.5) differs from the usual definition of the energy of point charges by the factor of 1/2.

Sometimes the factor is missing in Equation (4.5). In this case, the 1/2 shifted into the calculation equation of the electrostatic potential.

If one checks the literature, there is a lack of detailed explanation on the role of the 1/2 factors. One may have an intuition of it coming from the pairwise additive potential and/or of the constraint connecting the central and the image particles. If one compares Equations 4.1-4.4, it can be concluded that the 1/2 factor appears in the intra term, if one is interested in all particles within at least the cell. The 1/2 multiplier of the self-term is necessary, if one takes into account the periodic images. The 1/2 factor of the distinct term appears only if all particles in all cells are concerned.

At first, we checked the possible theoretical reasons and proofs on the energy partition, but none of them seemed to be helpful with complete certainty. Neither theoretician colleagues were able to provide simple convincing arguments. Therefore, we performed simple numerical experiments to choose the correct partition. It is known that Monte Carlo and molecular dynamic simulations provide the same results in accordance the equivalence of the time and ensemble averages. The force acting on a particle and on all of its images is straightforwardly given by the Newtonian laws, it is the appropriate derivative of Equation (4.2) divided by N. We performed molecular dynamics simulation where n=20-100 and the particles interact with Lennard-Jones potential. This potential is not a long-range interaction, but in the case of small number of particles, its effective range exceeds the size of the cell. Simultaneously, we performed Monte Carlo simulation on the same systems, where we used alternatively U^{tot}/N or U^{1} to calculate the potential energy. Our calculations showed numerically, that the U^{tot}/N definition is the correct one. The MC and the MD simulations provided the same structural results only in this case. The pair correlation functions of a molecular dynamics, a MC simulation with U^{tot}/N, and a MC one with U^{1} are shown in Figure 4.2. The energy conservation was satisfied also for the U^{tot}/N definition in the micro-canonical molecular dynamics simulation.

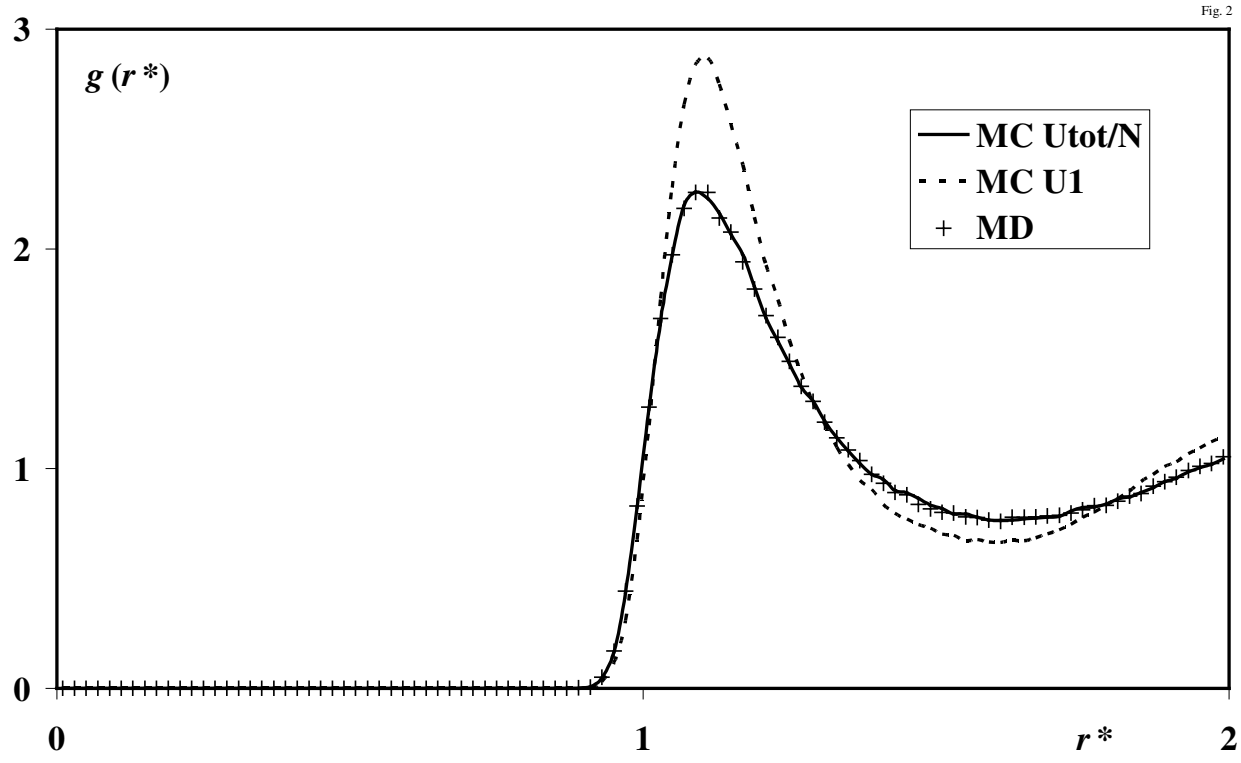


Figure 4.2 Pair correlation functions of Lennard-Jones systems to check the partition choice of potential energy. The system consisted of 40 Lennard-Jones particles close to the triple point. *r** is in reduced unit (Allen and Tildesley 1987).

4.3 Details of the Calculations

Let us recur to the main aim of our study: to the parameterisation of the Coulomb interaction. We describe the data sets, the formulae and the methods applied for the parameterisation in this section. We put two particles with opposite unit charges in the central cell of a system like in Figure 4.1. We centred the whole system on the position of the first particle. It meant we had one particle in the origin and another somewhere in the cell. We defined a Coulomb-like interaction $u_{ij}^{kl} = z_i z_j / r_{ij}^{kl}$, where r_{ij}^{kl} is the scalar distance of the *i*-th particle in the *k*-th cell and the *j*-th particle in the *l*-th one. For the sake of simplicity, we defined all variables to be dimensionless, e.g. they were normalised by the edge length of the cell and by the magnitude of the charges. In the cases of parameterisation and interpolation methods, the reduction of the space was feasible, wherein the process would be performed. One could use many simplifications for isotropic systems. If one puts the origin at the centre of the central cell, one could define eight

sub-cells, where only the sign of the second particle coordinates differed. The overall energy of the system was the same for all the eight positions, there were differences only in the sign of the forces. Furthermore, the choice of the axis was occasional. One might rotate the axis to define the coordinates of the second particle always to be $z \le y \le x$. It reduced the important space inside the cubic to its 1/6. If we used both the sign invariance and the choice of axis one, it was enough to fit or interpolate the interaction in a 1/48 part of the original cell. The shape of this reduced space was a tetrahedron inside the cubic central cell. Of course, if we used this reduced space in a real calculation, e.g. during a molecular simulation, we had to keep on file the original sign and order of the coordinates and we had to calculate the forces due to the original signs and axis.

The reference data were calculated both by the trivial method and by the Ewald one as it is implemented in (Allen and Tildesley 1987). The difference was rather small. It was below 10^{-4}%, if we chose the maximum of the inverse-space sums large enough and we used an appropriate width for the Gaussian-distributions in the Ewald method, and if we calculated the trivial sum over a large number of image cells. In the presented calculations we used 25 for the maximum of the square of the direct space sums and $L/5$ for the Gaussian width (see e.g. (Allen and Tildesley 1987; Frenkel and Smit 1996)), where L denotes the edge of the cell. (L=1 was used in all of the presented calculations). In the case of the trivial method, we calculated the interactions for particles in the central and other cells for all cases, where the centre of the image cell was closer to the central one than $100L$. It meant slightly more than four million image cells for each interaction. We should mention that a $25L$ or a $50L$ radius for the system supplied satisfactory results. We used $100L$ only to get very accurate data.

We calculated two types of data sets with random choice of the second particle position for the polynomial and rational function fits. 10000 random positions (vectors) were stored for the second particle in the first type, and the following inequalities were applied:

$$0 \le z \le y \le x \le 0.5 \qquad \textbf{(4.6)}$$

We calculated ten sets of these data. The second type of data sets was designed for the fit process. We intended to fit the potential energy by polynomials and rational functions, but we wanted to use the fitted functions for the calculation of both potential energies and forces. Since we used only the 1/48 part of the cell, we had to take care of the curvature of the fitted function at the surface of our elementary tetrahedron as well. The usual concept could be the incorporation of some constraints into the fit, like it is useful in the case of one-dimensional cubic splines and fast

interpolation methods (Press et al. 1992). Unfortunately, the situation was rather complicated in our three-dimensional case, and the non-violable constraints might reduce remarkably the degree of freedom in the polynomials. Therefore, we applied a different method to ensure smooth behaviour at the surfaces of the tetrahedron. We calculated data sets, where we allowed slight violence of the inequalities of Equation (4.6). We performed it by adding a random number of the [-0.05; 0.05] interval to the non-violating coordinates. Geometrically it meant an enlargement of the tetrahedron in all directions. We calculated two sets of 10000 particle coordinates with this second type of method. These sets were used only in the parameterisation procedures, while the first type of coordinates was used only in the test of the fitted functions.

We omitted the intra term ($\approx r_{ij}^{11}$ within the central cell) in the fit, because it can be calculated easily in an explicit way. We defined an interaction called as a correction term:

$$E_{ij}^{\mathrm{corr}} = z_i z_i \sum_{l\neq 1}^{N} 1/r_{ii}^{1l} + z_i z_j \sum_{l\neq 1}^{N}\sum_{j\neq i}^{2} 1/r_{ij}^{1l} = \sum_{l\neq 1}^{N} 1/r_{ii}^{1l} - \sum_{l\neq 1}^{N}\sum_{j\neq i}^{2} 1/r_{ij}^{1l} \qquad \textbf{(4.7)}$$

The intuitive description of the term, that it is the interaction between two particles: particle one in the central cell and particle one and two in the image cells, if n=2. The charges are omitted on the right hand side, because i and j were oppositely charged with unit absolute value in our two-particle systems. The total potential energy of one cell with arbitrary n can be calculated in a pairwise manner, in accordance with the previous section about the partition of the energy in periodic systems. Supposing that the net charge is zero in the central cell:

$$U^{\mathrm{tot}}/N = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} z_i z_j \left(1/r_{ij}^{11} - E_{ij}^{\mathrm{corr}}\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(u_{ij}^{11} - z_i z_j E_{ij}^{\mathrm{corr}}\right) \qquad \textbf{(4.8)}$$

The net zero charge is a necessary condition, otherwise U^{tot}/N contains incorrect number of self-term interactions.

The coefficients of the polynomials were determined in the usual way of ordinary least square fits on $E_j^{\mathrm{corr}}(x_j, y_{j,} z_j)$. The index of the first particle (i-th particle) was omitted, because this particle was in the origin of the two-particle periodic systems. The polynomials contained all possible terms up to a defined power:

$$P_M(x,y,z) = \sum\nolimits^{\mathrm{per}} a_{bcd} x^b y^c z^d \qquad \textbf{(4.9)}$$

The superscript per denotes all the possible terms with different *b, c* and *d*, where *b, c* and *d* are integers of [0;*M*], *M* is the highest power of the polynomial, and $(b+c+d)\leq M$. We fitted seven polynomials, $1\leq M \leq 7$.

In the case of the rational function fit, we used $R_{MS}(x,y,z)=\frac{Q_M(x,y,z)}{Q_S(x,y,z)}$, where the numerator and the denominator were similar polynomials as the ones defined in Equation (4.9), but $5\leq M+S\leq 7$. The optimal coefficients were determined by the method of Gauss-Newton-Marquardt (Press et al. 1992) as it is implemented in the Mathematica software (Wolfram 1999). In the test calculations the potential energies were calculated according to the fitted functions. The forces were calculated as the corresponding partial derivatives of the fitted functions.

In the case of the tabulated calculations, we defined a grid inside the tetrahedron described above. We used the 1/48 reduction of the central cell to spare the computational memory. The tetrahedron was enlarged to have at least two grid points in all the three Cartesian directions around the non-violating tetrahedron. We calculated the energies and also the *x, y* and *z* forces at all grid points. The intra term was subtracted from the data to be consistent to the polynomial and rational function fits. Since a cell with unit edge lengths was used, a bin size Δr=0.02 meant 25+4 grid points in all directions. It meant 29×30×31/6=4495 grid points, Δr =0.01 with 54 grids/direction counted 27720 points, and Δr=0.005 with 104 grids/direction gave 192920 data points. We applied two interpolation formulas in the test calculations. A tri-linear form (Wolfram 1999) was applied as

$$E_l^{\mathrm{corr}}=\sum_{i=I}^{I+1}\sum_{j=J}^{J+1}\sum_{k=K}^{K+1}\left(1-\left|x-x_i\right|\right)\left(1-\left|y-y_j\right|\right)\left(1-\left|z-z_k\right|\right)E_{ijk}^{\mathrm{corr}}/\Delta r^3 \,. \qquad \textbf{(4.10)}$$

x laid between x_I and x_{I+1}, *y* laid between y_J and y_{J+1}, *z* laid between z_K and z_{K+1}.

The other interpolation method was a tri-cubic one (Press et al. 1992). The interpolation was performed as one-dimensional cubic polynomial interpolation at first in the *x*, then in the *y*, and then in the *z* directions. One-dimensional cubic polynomials were fitted at first through the $\mathbf{r}_{I-1,j,k}$, $\mathbf{r}_{I,j,k}$, $\mathbf{r}_{I+1,j,k}$ and $\mathbf{r}_{I+2,j,k}$ points with $J-1\leq j\leq J+2$ and $K-1\leq k\leq K+2$, and the function values were interpolated to the $\mathbf{r}_{x,j,k}$ points. Thereafter it was repeated for the $\mathbf{r}_{x,J-1,k}$, $\mathbf{r}_{x,J,k}$, $\mathbf{r}_{x,J+1,k}$ and $\mathbf{r}_{x,J+2,k}$ points with $K-1\leq k\leq K+2$, and the function values were interpolated to the $\mathbf{r}_{x,y,k}$ points. Finally these last points were fitted by a cubic polynomial and the $\mathbf{r}_{x,y,z}$ point was interpolated. The polynomials were determined via the Lagrange method. It is rather expensive method of

calculation, and it is not recommended in time-consuming computer simulations, but as our results showed, the tri-cubic interpolation has very limited advantages anyway.

To compare our data with previous analytic functions fitted, we choose three functions of Adams and Dubey (Adams and Dubey 1987). The first one was an isotropic approximation, where the interaction over the $1/r$ minimal image convention term was described as

$$E_j^{\mathrm{corr}}(x_j, y_{j,} z_j) = 2.75022r^2 - 2.94414r^4 + 0.86910r^6 \qquad \textbf{(4.11)}$$

The other two equations based on a series extension of cubic harmonics (Equation 18. in (Adams and Dubey 1987)), where the potential energy is expanded as a finite sum of functions depending on r^l and $(x^l+y^l+z^l)$. Adams and Dubey determined different coefficients for sets including all even l powers up to given maxima of l values. They proposed theoretically derived and computationally optimised sets up to l=20. Unfortunately, there were typing errors in their tables (see also in Refs. (Ruff et al. 1989; Ruff and Diestler 1990)), therefore the l>14 sets did not provide satisfactory results. We used in our comparison the theoretical set with a maximal l=14 and an optimised set with maximal l=12.

4.4 Results and Discussions

The results of the test calculations on the 10 times 10000 data points are summarised in Tables 4.1-4.3. Table 4.1 contains the data on the fit of the potential to compare the polynomial fits of up to the 7$^{\mathrm{th}}$ degree, the rational function fits, where the sum of the degrees in the numerator and in the denominator were in the range of 5-7, data calculated with the aim of tri-linear and tri-cubic interpolations, and data obtained with previous analytical fits. The merit-function of the fits was defined as the average square difference of the original correction and the fitted one. One can see that the polynomial of the 7$^{\mathrm{th}}$ degree provided the best results, and the second best one was obtained by a rational function of 6/1 degrees. The performance of the tabulated methods was rather weak, they were slightly better than a polynomial of the second degree. The isotropic fit of Adams and Dubey performed weak. The optimised cubic harmonic fit (l≤12) provided better results than the cubic harmonic with theoretical parameters with more functions (l≤14). The optimised set provided accuracy around polynomials of the 4$^{\mathrm{th}}$ or 5$^{\mathrm{th}}$ degree. The trends of our fits are clear in this column of Table 4.1, the polynomials of high degree performed better. In the case of rational functions, the increasing sum of the degrees was the most important factor. If the sum was equal for two rational functions, the one with higher degree in the numerator provided better

data. The tabulated methods were sensitive primarily to the grid of the tabulation. The tri-cubic interpolation method performed slightly better than the simple tri-linear one.

Type		$\left\langle \left(E_i^{\text{fit/tab}} - E_i\right)^2 \right\rangle$	$100\left\langle \left\lvert \frac{E_i^{\text{fit/tab}} - E_i}{E_i} \right\rvert \right\rangle$	$100\left\langle \left\lvert \frac{E_i^{\text{fit/tab}} - E_i}{U_i} \right\rvert \right\rangle$	$\left\langle \left\lvert E_i^{\text{fit/tab}} - E_i \right\rvert \right\rangle$	$100\left\lvert \frac{\sum \left(E_i^{\text{fit/tab}} - E_i\right)}{\sum E_i} \right\rvert$
Polynomial fit	P-1	1.8E-03	1.5E+05	1.8E+00	3.2E-02	1.2E+00
	P-2	5.9E-05	6.7E+04	2.8E-01	5.6E-03	4.6E-01
	P-3	3.2E-06	4.2E+03	6.9E-02	1.3E-03	2.3E-02
	P-4	3.1E-07	2.3E+03	2.0E-02	4.1E-04	1.3E-02
	P-5	2.6E-08	1.7E+02	5.7E-03	1.1E-04	9.6E-03
	P-6	2.4E-09	1.1E+02	1.7E-03	3.4E-05	**7.3E-04**
	P-7	**2.3E-10**	1.1E+02	**5.3E-04**	**1.0E-05**	**6.2E-04**
Rational function fit	R-3/2	5.9E-07	4.0E+03	3.4E-02	5.7E-04	2.3E-01
	R-4/1	1.4E-07	4.1E+03	1.3E-02	2.7E-04	2.2E-02
	R-3/3	7.5E-08	3.7E+03	1.1E-02	2.2E-04	3.6E-02
	R-4/2	2.4E-08	1.1E+03	7.0E-03	1.3E-04	4.8E-03
	R-5/1	1.3E-08	1.8E+02	4.3E-03	8.1E-05	2.9E-03
	R-4/3	6.7E-09	6.3E+02	3.5E-03	6.5E-05	7.6E-03
	R-5/2	2.2E-09	1.2E+02	2.0E-03	3.6E-05	2.8E-03
	R-6/1	**8.0E-10**	2.3E+02	**1.0E-03**	**2.1E-05**	1.3E-03
Tabulated tri-linear	small	6.8E-05	6.2E+00	1.2E-01	1.4E-03	2.1E+00
	medium	3.3E-05	1.5E+00	6.0E-02	7.1E-04	1.1E+00
	large	1.7E-05	**6.5E-01**	3.0E-02	3.5E-04	5.3E-01
Tabulated tri-cubic	small	4.9E-05	4.7E+01	1.7E-01	2.0E-03	1.4E+00
	medium	2.5E-05	2.1E+00	9.1E-02	1.1E-03	7.7E-01
	large	1.3E-05	**7.0E-01**	4.8E-02	5.6E-04	4.0E-01
Isotr. fit	≤6	3.0E-03	2.6E+03	2.4E+00	4.5E-02	3.3E+01
Cubic harmonics	≤14	3.8E-06	1.2E-01	3.5E-02	3.1E-04	1.9E-01
	≤12opt.	1.1E-07	2.7E+00	1.1E-02	1.7E-04	1.1E-02

Table 4.1 Comparison of potential energy for the polynomial fits, rational function fits, tabulated methods and cubic harmonic fits. For statistical and other details see the text.

At the creation of tabulated data, the average square difference was not a criterion as in the parameterisation methods. Therefore, it is necessary to use other measures in the comparison of

the different methods. The next three columns contain further data. Surprisingly, the tabulated method performed with three magnitudes better, if we calculate the averaged relative difference of the fitted and original corrections: $100\left\langle\left|\frac{E_i^{\text{fit/tab}}-E_i}{E_i}\right|\right\rangle$. How is this reverse order possible? How can be this relative average error of the fits around and over 100%? We checked the worst data and we found that there are some points in each data set, where the two particles are very close to each other. In these cases, the correction terms are very small and the intra part of the energy can be about 10 magnitudes larger than E_i-s. For example, in the worst data of the polynomial fit of 7th degree U_i was around -170, E_i was -1×10^{-9}, and E_i^{fit} was 7×10^{-5}. It means a 7,000,000% relative error at this point, but in the absolute scale, the error is only around 10^{-5}, or relatively to the intra term, it is only 0.00004%. The polynomial and the rational function fits performed poorly in this relative comparison, while the tabulation works stable. The cubic harmonic methods were better in this comparison, than the polynomials. If we use the total interaction in the normalisation, $100\left\langle\left|\frac{E_i^{\text{fit/tab}}-E_i}{U_i}\right|\right\rangle$, the fit methods provides about two magnitudes better results than the tabulated ones. If this criterion was used, the worst data represented the cases, where the intra terms were small, the second particles were close to the corner of the cell, and the correction term was very important. The best average error was 0.0005% in the fits and 0.05% in the tabulation in this comparison. If we simply compared $\left\langle\left|E_i^{\text{fit/tab}}-E_i\right|\right\rangle$, the fits of high degrees performed one magnitude better than the tabulation. The calculation of potential energy has a crucial role in the Monte Carlo simulation methods. Here, the total energy of the system is important. It means, the fitting and the tabulated methods should work correctly not only for one interaction, but it is necessary to provide good total energies. Each of our test sets contained 10000 data. The number of these interactions corresponds to a system about 140 charged particles. The last column in Table 4.1 shows the average in the 10 sets of $100\left|\frac{\sum\left(E_i^{\text{fit/tab}}-E_i\right)}{\sum E_i}\right|$. One can see, the best fits gave the overall corrections with 0.0006% error. The tabulation provides 0.4% error. We concluded that the polynomial and rational function parameterisations of high degree provided expressively better results than the tabulation. The tabulation was superior only in one relative comparison, but here the defective points of the fits were those ones,

where the corrections were very small and unimportant on the absolute scale. The performance of the optimised cubic harmonic fit in the last three comparisons was similarly around the polynomials of 4th and 5th degree.

The two most important methods in the classical mechanical modelling of liquid matter are the Monte Carlo and the molecular dynamic simulations. In the case of Monte Carlo methods, the calculation of the potential energy is necessary, so the above fits can be important. In the case of molecular dynamics, the calculation of forces is essential. We present the same table for the forces as Table 4.1 for the potential. The forces were calculated simply as the corresponding *x*, *y* and *z* partial derivatives of the fitted potential. In the case of the tabulation, the forces were tabulated similarly to the potential, and the same tri-linear or tri-cubic interpolation methods were applied. F_i denotes the original correction term, $F_i^{\text{fit/tab}}$ is the fitted or tabulated value of the correction term, and F_i^{all} means the sum of the correction and the intra term. The data shown in Table 4.2 are the averages over the ten test sets and the three directions: *x*, *y*, and *z*. Trends were similar to the ones in Table 4.1. The fits with polynomials or rational functions of high degree performed the best. In the case of the relative comparison to the force correction, the tri-linear seems to be superior, but the apparently weak performance of the fits did not mean serious deficiency in this comparison. We checked the worst data, and we concluded similarly as in the case of the potential. The weak average originates from the points, where the contribution to the force from the intra part is about 10 magnitudes larger than the correction term. The fits provide maximum about four magnitudes larger corrections, than the real ones, but these ones are still negligible in respect to the total force. We omitted the line, which corresponds to the total force within one data set, because it is not important to calculate it in molecular dynamics. The optimised cubic harmonic sets performed again better than the theoretical sets. Its accuracy laid between the polynomials of 4^{th} and 5^{th} degree for the first comparison, but in the third and fourth comparison they provided slightly better data than the polynomials of 5^{th} degree.

We did not provide confidence intervals for the data in Table 4.1 and Table 4.2, because we did not want to overcrowd the tables. We calculated the error intervals for the data with a significance of α=0.05 on the 10 test sets. For the force data it meant 30 sets taking an average over the *x*, *y*, and *z* directions. The relative uncertainty of the data was slightly different for the various columns in the two tables. The average error was between 0.02-0.06% for the first, third and 4th columns of Table 4.1 in the cases of polynomial and rational function fits. Respectively,

it was in the 0.05-0.2% range for the force data in Table 4.2. The corresponding values for the tabulated data were 2-5 times larger. In the case of the second columns, the error bars were smaller for the tabulation (for the energy around 1.5% and 0.2% for the forces), and the data for the fits were about 3 times larger. The uncertainty of the 5th column in Table 4.1 was 0.8% for the fits and 0.2% for the tabulations.

Type		$\left\langle \left(F_i^{\text{fit/tab}} - F_i\right)^2 \right\rangle$	$100\left\langle \left\lvert \frac{F_i^{\text{fit/tab}} - F_i}{F_i} \right\rvert \right\rangle$	$100\left\langle \left\lvert \frac{F_i^{\text{fit/tab}} - F_i}{F_i^{\text{all}}} \right\rvert \right\rangle$	$\left\langle \left\lvert F_i^{\text{fit/tab}} - F_i \right\rvert \right\rangle$
Polynomial fit	P-2	1.9E-02	1.8E+03	1.5E+01	9.6E-02
	P-3	1.5E-03	2.7E+02	5.2E+00	2.4E-02
	P-4	1.9E-04	2.6E+02	1.4E+00	8.4E-03
	P-5	1.5E-05	6.3E+01	4.6E-01	2.4E-03
	P-6	1.4E-06	1.6E+01	1.1E-01	7.7E-04
	P-7	**1.8E-07**	4.1E+00	**3.9E-02**	**2.6E-04**
Rational function fit	R-3/2	3.1E-04	1.1E+02	2.6E+00	1.1E-02
	R-4/1	6.0E-05	1.9E+02	8.5E-01	5.2E-03
	R-3/3	3.5E-05	5.5E+01	9.4E-01	4.2E-03
	R-4/2	1.2E-05	5.9E+01	4.9E-01	2.4E-03
	R-5/1	7.8E-06	4.0E+01	3.4E-01	1.7E-03
	R-4/3	3.8E-06	4.3E+01	2.5E-01	1.3E-03
	R-5/2	1.4E-06	2.3E+01	1.9E-01	7.3E-04
	R-6/1	**4.1E-07**	1.6E+01	**6.1E-02**	**4.4E-04**
Tabulated tri-linear	small	7.2E-04	2.4E+00	2.7E-01	4.7E-03
	medium	3.5E-04	9.5E-01	1.2E-01	2.3E-03
	large	1.8E-04	**5.0E-01**	5.7E-02	1.1E-03
Tabulated tri-cubic	small	5.5E-04	3.2E+00	4.0E-01	6.9E-03
	medium	2.7E-04	1.5E+00	1.8E-01	3.6E-03
	large	1.4E-04	**6.6E-01**	9.4E-02	1.8E-03
Isotr. fit	≤6	2.4E-01	4.7E+02	1.5E+01	3.4E-01
Cubic harmonics	≤14	6.5E-04	9.5E-01	2.4E-01	4.7E-03
	≤12opt.	3.0E-05	1.0E+00	1.4E-01	2.5E-03

Table 4.2 Comparison of forces for the polynomial fits, rational function fits, tabulated methods and cubic harmonic fits. For statistical and other details see the text.

If one looks at Table 4.1 and Table 4.2, the optimised solution would be a polynomial fit of high degree. Of course, in the case of computational methods, the requested number of operations is an important factor. We summarised the number of different operations for the calculation of potential energy and force in Table 4.3. The data were calculated in respect to one interaction. The calculation cost of the distance of the two interacting particle is not included in the table. In the case of the polynomial and rational function fits and the tabulation, the calculation cost of the bookkeeping of the original signs of the coordinates and of the rotation of the vector to provide $0 \le z \le y \le x \le 0.5$ was not counted. The table contains data on the operations demanded by the potential calculation via the Ewald method as it was implemented in the supplement of the book of Allen and Tildesley (Allen and Tildesley 1987). If one reports on the scaling of the Ewald method, there are different data from n to n^2, both for the direct and the reciprocal space sums. The difference in the data originates from the adjustable width of the Gaussian distribution that is a free parameter in the Ewald method. If the Gaussians are chosen to vanish at a cut-off of half the cell size, the direct sum scales n^2 and the reciprocal sum scales n. If the Gaussian width is chosen to provide Gaussian vanishing at a standard cut-off distance independent of n, the direct sum scales n, but the reciprocal sum becomes n^2 scaling (Essmann et al. 1995; Sagui and Darden 1999). By varying the cut-off with the square root of the cell size, it can be shown that both the direct and reciprocal sum can be scaled $n^{3/2}$, which is supposed to be optimal by some authors (Perram et al. 1988). As we mentioned in the introduction, the primary aim of our investigation was to get a practically usable interpretation of the periodic Coulomb interactions in the case of program codes, where the interactions are calculated via one by one. In the general implementation of the Ewald method, the calculation is performed on the interactions via one by one for the direct sum, but the reciprocal space sum is calculated only once for all of the interactions in the system. Therefore, if we would like to compare our methods to the Ewald one, we have to define a system size (n) and also the choice of the Gaussian width. The data in Table 4.3 correspond to a linear scaling in the reciprocal space sum and an n^2 in the direct part. The system sizes are indicated in the table, too. The unit of the number of the operations corresponds to one multiplication or addition. The other operations, as division, calculation of exponentials, trigonometric functions and logic functions were counted in these units. The ratios were obtained in small test calculations. In these tests we calculated the cost of the different functions by small programs, where the CPU time of a million of these calculations were counted relatively to a

million of multiplications or additions. The codes were written in Fortran and C, and they were used on PC-s with Pentium IV processor running under Windows XP. We found no relevant differences in the ratios in the cases of the two programming languages. Maybe it is a consequence of that we used the compilers from the same vendor for the two languages. There were not any significant differences in the calculation cost of a double precision addition and multiplication. An exponential took 22.9, a sine or cosine took 11.4, a division took 2.6 and a conditional jump took 1.6 times longer time, than a multiplication, addition or subtraction. The truncation of a real number to an integer cost 1.6 units. A search in a vector depended on the size, for vectors up to a few ten thousands elements it was three times more expensive than a multiplication, but it increased up to 13 times more for a vector with 192920 elements.

The corresponding column of Table 4.3 contains the multiplication equivalent calculation cost for the different methods. In the case of polynomials, the calculation needs solely multiplications and additions in a half-half ratio, if one uses the computationally feasible Horner arrangements. In the case of the rational functions, both the numerator and the denominator can be expressed in the Horner form. It was rather surprising, that the number of the operations were rather high already for the simple tri-linear interpolation. In the case of the tri-cubic interpolation, the operational cost was extremely high. It was coded with the expensive one-dimensional Lagrange method, but significant reduction is not possible with any other cubic methods.

If one compares the data of Table 4.1 and Table 4.3, we can find optimal solutions for the potential calculations. Let us compare the polynomial, the rational function, the tabulated methods and the cubic harmonics at first. The tabulated ones performed weakly with respect to the computational cost. Therefore, we do not suggest the use of tabulated methods. It is not easy to choose from the polynomials and rational functions of different degrees. Depending on the desired accuracy, we suggest the following order: P-3 < R-3/2 < P-4 < R-3/3 < R-4/3 < R-5/2 < R-6/1 < P-7. The notations correspond to the acronyms in the tables. We omitted some of the fits, where other fits with similar computational cost were close to them with higher accuracy. The number of the operations with respect to the merit function is visualised for the parameterised functions in Figure 3. The computational cost of the optimised cubic harmonics ($l \leq 12$) was slightly larger than the polynomial of 5th degree, but relevant difference was not found. In comparison with the rational functions the use of cubic harmonics seemed to be less feasible.

Type		Number of instructions	
		potential	forces and potential
Polynomial fit	P-1	6	-
	P-2	18	30
	P-3	38	74
	P-4	68	144
	P-5	110	246
	P-6	166	386
	P-7	238	570
Rational function fit	R-3/2	59	120
	R-4/1	77	166
	R-3/3	79	164
	R-4/2	89	190
	R-5/1	119	268
	R-4/3	109	234
	R-5/2	131	292
	R-6/1	175	408
Tabulated tri-linear	small	215	431
	medium	215	431
	large	295	751
Tabulated tri-cubic	small	2615	6555
	medium	2615	6555
	large	3255	9115
Ewald	n=10	710	2724
	n=50	143	514
	n=100	93	278
	n=10000	42	65
Isotr. fit	≤6	10	22
Cubic harmonics	≤14	145	629
	≤12opt.	121	537

Table 4.3 Number of instructions for the calculation of one interaction. The unit of the operations is one multiplication or addition.

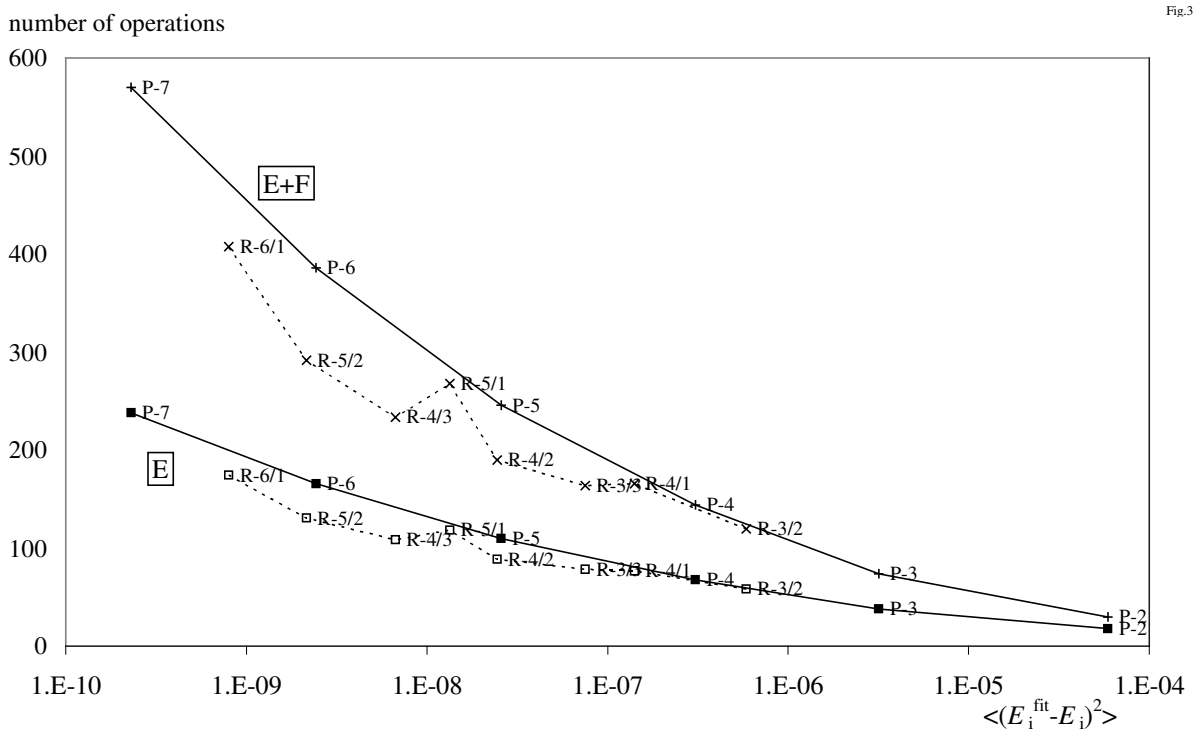


Figure 4.3 Number of operations with respect to the merit function of the fits (second column of Table 4.1). Energy calculations: □ polynomials, ■ rational functions. Energy and force calculations: + polynomials, × rational functions. The data are connected with solid and dashed lines to clarify trends.

If one compares the performance of the fits to the Ewald method, one can find optimal solutions with respect to the system size. A polynomial of the third degree is less expensive computationally than the Ewald method at all system sizes. An accurate fit costs more than the Ewald method for large systems. At the system size around 100 particles, or more explicitly, where the number of the charged particles is around 100, the rational functions with a degree of 4/3 or 3/3 are computationally comparable to the Ewald method. For small systems (less than 50 charged particles) the use of whatever fit is feasible.

The situation is different, if one is interested in the forces, too. Here one has to calculate 4 values: the *x*, *y* and *z* components of the forces and the potential. The estimated total computational costs of these calculations are summarised in Table 4.3. The data corresponding to the fits are approximate. There are many common terms in the Horner expressions of the potential and *x*, *y*, and *z* components of the forces in the polynomials and rational functions. Furthermore, the degree of the functions is one less in the force calculation than in the potential calculation. An

effective compiler may omit the redundant calculations. We guessed that the cost of these calculations might be reduced to the calculation of two forces and the potential instead of three forces and the potential. The data in Table 4.3 include this reduction also for the cubic harmonics. As one can see in Table 4.3, the force calculation with the cubic harmonics is expensive. The used degree of *r, x, y,* and *z* is rather large in the cubic harmonic expansion, and the loss of degree is less significant in comparison to the rational functions or polynomials. The calculation of a cubic harmonic force component cost more than a calculation of a potential value, while the opposite case happened for our polynomials and rational functions. The calculation cost of the optimized $l \leq 12$ cubic harmonics is around a polynomial with a degree of 7th, but the accuracy of the latter is reasonably better. The same accuracy can be achieved with a half of the computation time with polynomials and especially with rational functions, if both potential energy and forces are calculated. Therefore, it is feasible to use our new functions especially in molecular dynamics and in other applications, where the forces are important as well. The common calculation of forces and potential slightly modified the preferences in respect to the Ewald method. The polynomial of third degree became slightly less effective than the Ewald method for large systems. On contrary, for systems consisting about 100 charged molecules one can efficiently use more precise polynomials than in the case of simple potential calculations. The rational function 6/1 fit seemed to be the best choice for systems consisting of less than 50 particles (see also in Figure 4.3).

4.5 Conclusions

The aim of our study was to elaborate a simple method to calculate Coulomb interactions in three-dimensional periodic systems. The method is not intended to be competitive with well-established methods on physical intuition. We would like to present a feasible way to calculate the long-range interactions in program codes primarily designed for systems with short-range interactions. We have shown that the crucial part of the Coulomb interaction, which is the extra term besides the contribution within the minimal image convention, can be easily parameterised with simple functions. We used polynomials of up to the 7^{th} degree and rational functions with a sum of 5, 6 and 7 of the degrees in the numerators and denominators. The comparison of the fits to simple tabulated functions showed, that the accuracy of the fits is reasonably better than for the tabulations, especially if we took into account the computational cost of the methods. Our

functions seem to be as feasible as the previous cubic harmonic parameterisations, if solely the potential energy is calculated. The comparison was performed both for simple potential calculation and for common calculation of forces and potentials. The first corresponded to the MC simulations, while the second concerned the requirements of MD. If both potential energy and forces are needed, the polynomials and rational functions provide the same accuracy with a half of computational cost than it is required for cubic harmonics. We proposed some sets of parameterised functions. The suggested functions are detailed in the Appendix.

Of course, our parameterised functions should be competitive not only during the program code development, but they should not mean a drastic increase in the computational cost during the simulations. Therefore, we compared the calculation cost of our methods to the Ewald one. We obtained reasonable good performance for our methods in the case of small systems. In the case of medium and large systems, our parameterised polynomials and rational functions and the previously fitted cubic harmonics are competitive in the calculation cost, only if one uses less accurate fits.

The methods are intended to be easily embeddable into existing codes. One needs to add them as a function into the codes and call them at each interaction pair.

5 Summary

The inverse theorem of liquids states that there is a one to one correspondence between the structure of the liquid and the interaction among the particles. The practical solution of the inverse theorem in one direction is resolved: one can get the structure by molecular simulations for known interactions. In the other direction, there are no widely accepted methods up to now. We examined the power of the artificial neural networks as non-linear approximative methods for the solution of the inverse theorem in this structure to potential direction. We used a scheme of the traditional supervised trained, monolayer neural network. The input data were molecular dynamics simulation generated structure-factor pair-potential pairs, three types of pair potentials were used: Lennard-Jones, Morse and Buckingham potentials. We used altogether 2000 pairs for the training of the network and we tested the performance on further 200 function pairs. For the optimised solution, we checked many factors e.g. the number of hidden neurons, modifications of the merit functions, etc. An encouraging reproduction of the pair potentials has been attained, especially in the case of Lennard-Jones and Morse potentials.

In the second part of the thesis we focused on a computational challenge, namely the simulation of asymmetric binary hard-sphere systems as those systems are of great scientific interest and there was no comprehensive study especially regarding pair-correlation functions with large size ratios to be found in the literature. It is a cumbersome task according to the number of particles increasing cubic with the asymmetric ratio. We started with Monte Carlo technique, but for feasible results, we had to turn to molecular dynamics. We developed an optimised molecular dynamics code for 1:5 and 1:10 ratios considering the architecture of nowadays-available personal computers. The developed algorithm uses the link-cell method and the traditional calculation of next collision times, but according to the size asymmetry of the constituent particles, it was differentially interpreted for the particle types. Our performed simulations provided a good base to test different theories and approximations on binary hard sphere systems. The asymmetric ratio was 1:5 and 1:10 and the partial packing fractions of the spheres ranged from 0 to 0.5 in 0.1 increments as feasible. We obtained a reasonable set of partial pair correlation functions for these systems, in contrary to the majority of earlier studies also for the large-large partials. Our data provided a good basis for the test of the applicability of the one

component hard sphere approximation and for comparison to theoretical methods. For the limit of the approximation of two-component hard-sphere structure with one-component hard-sphere one: we found that this approximation cannot be applied, if the packing fraction of the large spheres is greater than about 0.2. We studied the performance of the Percus-Yevick and the Rational Function Approximation theoretical methods in the creation of pair correlation functions. The performances of these theoretical methods were convincing in the small-small and small-large cases. On contrary, the large-large pair correlation functions can be used only as a very crude approximation at medium and large densities. The two methods behave rather similarly, but the Percus-Yevick method seemed to be worse than the Rational Function Approximation method. In the comparison of the 1:5 and 1:10 data, we concluded that both the approximation of the one-component hard-sphere model and the theoretical models are less successful as we approach the colloid limit. Finally, we reported a set of contact values of $g(r)$-s for the binary hard-sphere systems studied by molecular dynamic simulations.

The simulation of systems with short range potentials can be solved with linear scaling algorithms while the presence of long range interactions increase the scaling up to n squared or even higher. There are plenty of more efficient solutions for different, mostly Coulomb systems. Some of the solutions involve a rather complicated program code or their applicability is limited to e.g. special system sizes. We developed a simple parameterisation of the Coulomb force, which can be easily embedded in codes written for short-range interactions. We started with a theoretical problem: how one should partition the energy of a periodic system. We showed with numerical calculations, using the ergodic equivalence of Monte Carlo and molecular dynamics simulations, that the most commonly used definition is the unique correct one. We decomposed the Coulomb interactions of the periodic systems into an intra-cell and two inter-cell parts. We randomly generated several sets number of points, each system consisted 10,000 distances with interactions and forces calculated by the Ewald and the trivial sum methods. The inter-cell interactions were parameterised by simple polynomials and rational functions. We used polynomials of up to the 7^{th} degree and rational functions with a sum of 5, 6 and 7 of the degrees in the numerators and denominators. We compared the performance of this parameterisations and earlier methods in detail.

Brief Summary

This work concentrates on the structure of disordered systems with the help of classical simulations; the articles on which this work is based are not tightly related.

Neureal networks have been successfully applied to solve the inverse problem of statistical mechanics: a neural network trained on a training set consisting of 2000 simulation generated pair potential – structure factor pairs provided an encouraging reproduction of the pair potentials, especially in the case of Lennard-Jones and Morse potentials.

A whole set of pair correlation functions of binary hard-sphere mixtures at 1:5 and 1:10 size ratios for a wide and systematic range of packing fractions have been obtained. For the limit of the approximation of two-component hard-sphere structure with one-component hard-sphere one: we found that this approximation cannot be applied, if the packing fraction of the large spheres is greater than about 0.2. We studied the performance of the Percus-Yevick and the Rational Function Approximation theoretical methods in the creation of pair correlation functions. These two methods were chosen only from a practical viewpoint. The performances of these theoretical methods were convincing in the small-small and small-large cases. On contrary, the large-large pair correlation functions lose their physical reality at medium and large densities. The two methods behave rather similarly, but the Percus-Yevick method seemed to be worse than the Rational Function Approximation method. The large-large partials of these analytical functions failed for dense systems. In the comparison of the 1:5 and 1:10 data, we concluded that both the approximation of the one-component hard-sphere model and the theoretical models are less successful as we approach the colloid limit. Finally, we reported a set of contact values of $g(r)$-s for the binary hard-sphere systems studied by molecular dynamic simulations.

A simple method to calculate Coulomb interactions in three-dimensional periodic systems has been devised. We have shown that the crucial part of the Coulomb interaction, which is the extra term besides the contribution within the minimal image convention, can be easily parameterised with simple functions. We used polynomials of up to the 7th degree and rational functions with a sum of 5, 6 and 7 of the degrees in the numerators and denominators.

Összefoglalás

A doktori dolgozat a rendezetlen rendszerek szerkezetével foglalkozik klasszikus számítógépes szimulációk segítségével. A munka alapját képező közlemények szorosan nem függenek össze.

Neurális hálózatokat sikeresen alkalmaztunk a statisztikus mechanikai inverz probléma megoldására: egy 2000 szimulációval generált párpotenciál – szerkezeti függvény páron betanított neurális hálózat a párpotenciálokat kielégítő módon visszaadta, különösen a Lennard-Jones és a Morse potenciálok esetében.

1:5 és 1:10 méretarányú kétkomponensű merevgömbi rendszerek párkorrelációs függvényeit egy széles és szisztematikusan felosztott térkitöltési hányadosú tartományban számoltuk számítógépes szimuláció segítségével. A kétkomponensű merevgömbi szerkezet egykomponensű merevgömbi rendszerrel való közelíthetőségéről az alábbiakat találtuk: ez a közelítés nem alkalmazható, ha a nagy gömbök térkitöltési hányadosa kb. 0,2-nél nagyobb. Tanulmányoztuk a Percus-Yevick és Rational Function Approximation elméleti közelítő módszerek teljesítményét párkorrelációs függvények előállítására. Ez a két módszer praktikus okokból lett választva. Ezen módszerek teljesítménye meggyőző a kis-kis és kis-nagy parciális szerkezetek esetében. Ezzel szemben a nagy-nagy párkorrelációs függvények elveszítik a fizikai realitásukat közepes és nagy sűrűségeknél. A két módszer meglehetősen hasonlóan viselkedik, a Percus-Yevick eljárás teljesítőképessége kisebb, mint az Rational Function Approximation módszeré. Mindkét elméleti közelítés rosszul becsüli a nagy-nagy parciális párkorrelációs függvényeket sűrűbb rendszerek esetén. Az 1:5 és 1:10 méretarányú adatokat összehasonlítva kimondhatjuk, hogy mind az elméleti közelítések, mind az egykomponensű merevgömbi modell egyre kevésbé alkalmazható, ahogy közelítünk a kolloid-oldószer méretarány felé. A kutatás egyik eredményének tekinthetők a párkorrelációs függvény kontaktértékei is.

Egy egyszerű eljárást fejlesztettünk ki Coulomb kölcsönhatások háromdimenziós rendszerekben való számolására. Megmutattuk, hogy a Coulomb kölcsönhatás döntő része, ami a ‘minimal image convention’-on felüli többlet tag, könnyedén parametrizálható egyszerű függvényekkel: hetedik rendig felmenő polinomokat és racionális törtfüggvényeket, amelyekben a számláló és a nevező rendjének összege 5, 6 és 7 volt, használtunk.

Bibliography

http://www.mathematica.com.

Adams, D. J. and Dubey, G. S. (1987). J. Comp. Phys. **72**: 156.

Alder, B. J. and Wainwright, T. E. (1956). Molecular dynamics by electronic computers. Int. Symp. on Statistical Mechanical Theory of Transport Processes Brussels, Interscience, Wiley, New York.

Allen, M. P. and Tildesley, D. (1987). Computer Simulation of Liquids. Oxford, Oxford University Press.

Almarza, N. G. and Lomba, E. (2003). Phys. Rev. E **68**: 011202

Almarza, N. G. and Lomba, E. (2004). Phys. Rev. E **70**: 021203

Amokrane, S., Ayadim, A. and Malherbe, J. G. (2005). Journal of Chemical Physics **123**(17): -.

Anderson, F. G., Leung, C.-H. and Ham, F. S. (2000). J. Phys. A: Math. Gen. **33**: 6889.

Asakura, S. and Oosawa, F. (1958). Journal of Polymer Science **33**: 183.

Attard, P. (1989). Journal of Chemical Physics **91**(5): 3083-3089.

Baranyai, A. and Schiller, R. (2003). Statisztikus mechanika vegyészeknek. Budapest, Akadémiai.

Barrio, C. and Solana, J. R. (2000). Journal of Chemical Physics **113**(22): 10180-10185.

Bergmann, A., Fritz, G. and Glatter, O. (2000). Journal of Applied Crystallography **33**: 1212-1216.

Biben, T. and Hansen, J. P. (1991). Physical Review Letters **66**(17): 2215-2218.

Borbely, S. (2000). Langmuir **16**(13): 5540-5545.

Boublik, T. (1970). Journal of Chemical Physics **54**: 1523.

Bourasseau, E., Haboudou, M., Boutin, A., Fuchs, A. H. and Ungerer, P. (2003). J. Chem. Phys. **118**: 3020

BrunnerPopela, J. and Glatter, O. (1997). Journal of Applied Crystallography **30**: 431-442.

Brush, S. G., Sahlin, H. L. and Teller, E. (1966). J. Chem. Phys. **45**: 2102.

Cao, D. P., Chan, K. Y., Henderson, D. and Wang, W. C. (2000). Molecular Physics **98**(9): 619-624.

Cho, K. H., No, K. T. and Scheraga, H. A. (2002). J. Mol. Struct. **641**: 77.

Clement-Cottuz, J., Amokrane, S. and Regnaut, C. (2000). Physical Review E **61**(2): 1692-1702.

Darden, T. A., York, D. and L., P. (1993). J. Chem. Phys. **98**: 10089.

Dickman, R., Attard, P. and Simonian, V. (1997). Journal of Chemical Physics **107**(1): 205-213.

Dijkstra, M., van Roij, R. and Evans, R. (1999). Physical Review E **59**(5): 5744-5771.

Eastwood, J. W., Hockney, R. W. and Lawrence, D. (1980). **19**: 215.

Esselink, K. (1995). Comp. Phys. Comm. **87**: 375.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T. A., Lee, H. and Pedersen, L. (1995). J. Chem. Phys. **103**: 8577.

Ewald, E. E. (1921). Ann. Phys **64**: 253.

Foffi, G., Gotze, W., Sciortino, F., Tartaglia, P. and Voigtmann, T. (2003). Physical Review Letters **91**(8): -.

Frenkel, D. and Smit, B. (1996). Understanding molecular simulation London, Academic.

Friedman, H. L. (1975). Mol. Phys **29**(1533).

Fries, P. H. and Hansen, J. P. (1983). Molecular Physics **48**(5): 891-901.

Gassner, H., Probst, M., Lauenstein, A. and Hermansson, K. (1998). J. Phys. Chem. A **102**: 4596.

Germain, P., Regnaut, C. and Amokrane, S. (2003). Physical Review E **67**(6): -.

Greengard, L. and Rokhlin., V. (1987). J. Comp. Phys. **73**: 325.

Haberlandt, R., Fritzsche, S., Peinel, G. and Heinzinger, K. (1995). Molekular-dynamik. Wiesbaden, Vieweg.

Habershon, S., Cheung, E. Y., Harris, K. D. M. and Johnston, R. L. (2004). J. Phys. Chem. A **108**: 711.

Hansen, J. P. (1973). Phys. Rev. A **8**: 3096.

Hansen, J. P. and McDonald, I. R. (1986). Theory of simple liquids. London, Academic.

Henderson, D., Malijevsky, A., Labik, S. and Chan, K. Y. (1996). Molecular Physics **87**(2): 273-285.

Henderson, R. L. (1974). Phys. Lett. A **49**: 197.

Heno, Y. and Regnaut, C. (1991). Journal of Chemical Physics **95**(12): 9204-9208.

http://www.nd.com.

Hu, L. H., Wang, X. J., Wong, L. H. and Chen, G. H. (2003). J. Chem. Phys. **119**: 11501.

Imhof, A. and Dhont, J. K. G. (1995). Physical Review Letters **75**(8): 1662-1665.

Isobe, M. (1999). International Journal of Modern Physics C **10**: 1281.

Jesús, O. D. and Hagan, M. T. (2007). IEEE Transactions on Neural Networks **18**.

Johnson, M. D., Hutchinson, P. and March, N. H. (1964). Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences **282**: 283-302.

Jorgensen, W. L., Madura, J. D. and Swenson, C. J. (1984). J. Am. Chem. Soc. **106**: 6638-6646.

Kahl, G. and Pastore, G. (1991). Journal of Physics a-Mathematical and General **24**(13): 2995-3011.

Kiss, I. Z., Mandi, G. and Beck, M. T. (2000). J. Phys. Chem. A **104**: 10994.

Krantz, A. T. (1993). ACM Trans. Mod. Comp. Sim **6**: 185.

Ladd, A. J. C. (1977). Mol. Phys **33**: 1039.

Ladd, A. J. C. (1978). Mol. Phys. **36**: 463.

Larsen, B., Førland, T. and Singer, K. (1973). Mol. Phys **26**: 1521 - 1532.

Lebowitz, J. L. (1964). Physical Review A **133**: 895.

Lebowitz, J. L. and Rowlinson, J. S. (1964). Journal of Chemical Physics **41**: 133.

Lilichenko, M. and Kelley, A. M. (2001). J. Chem. Phys. **114**: 7094.

Luty, B. A., Tironi, I. G. and van Gunsteren, W. E. (1995). J. Chem. Phys. **103**: 3014.

Lyubartsev, A. P. and Laaksoonen, A. (1995). Phys. Rev. E **52**: 3730

Malherbe, J. G. and Amokrane, S. (2001). Molecular Physics **99**(4): 355-361.

Malijevsky, A., Barosova, M. and Smith, W. R. (1997). Molecular Physics **91**(1): 65-73.

Malijevsky, A. and Veverka, J. (1999). Physical Chemistry Chemical Physics **1**(18): 4267-4270.

Mansoori, A., Carnahan, N. F. and Starling, K. E. (1971). Journal of Chemical Physics **54**: 1523.

Marin, M., Risso, D. and Cordero, P. (1993). Journal of Computational Physics **109**(2): 306-317.

McGreevy, R. L. and Pusztai, L. (1988). Mol. Sim. **1**: 369.

Metropolis N, Rosenbluth AE, Rosenbluth MN, Teller AH and E, T. (1953). J Chem Phys **21**: 1087.

Momany, F. A., Carruthers, L. M., Macguire, R. F. and Scheraga, H. A. (1974). J. of Phys. Chem. **78**: 1595.

Neumann, M. and Steinhauser, O. (1980). Mol. Phys. **39**(437).

Panek, J. J., Jezierska, A. and Vracko, M. (2005). J. Chem. Inf. and Mod **45**: 262.

Paschinger, E., Reiner, A. and Kahl, G. (1998). Molecular Physics **94**(5): 743-750.

Pedersen, J. S. (1994). Journal of Applied Crystallography **27**: 595-608.

Perram, J. W., Petersen, H. G. and DeLeeuw, S. W. (1988). Mol. Phys. **65**: 875

Pollock, E. L. and Glosli, J. (1996). Comp. Phys. Comm. **95**: 93.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). Numerical recipes in Fortran. Cambridge, Cambridge University Press.

Preu, H., Zradba, A., Rast, S., Kunz, W., Hardy, E. H. and Zeidler, M. D. (1999). Physical Chemistry Chemical Physics **1**(14): 3321-3329.

Pusztai, L., Gereben, O. and Baranyai, A. (1995). Physica Scripta **57**: 69.

Raff, L. M., Malshe, M., Hagan, M., Doughan, D. I., Rockley, M. G. and Komanduri, R. (2005). J. Chem. Phys. **122**: 84104.

Ramsay, J. D. F. and Lindner, P. (1993). J. Chem. Soc., Faraday Trans. **89**: 4207-4214.

Rapaport, D. C. (1995). The art of molecular dynamics simulation. Cambridge, Cambridge University Press.

Reatto, L. (1982). Phys.Rev. B **26**: 130

Reatto, L. (1988). Phil. Mag. A **58**: 37.

Rosenfeld, Y. (1994). Physical Review Letters **72**(24): 3831-3834.

Rosenfeld, Y. and Aschroft, N. W. (1979). Phys. Rev. A **20**: 1208 - 1235.

Roth, R., Evans, R. and Dietrich, S. (2000). Physical Review E **62**(4): 5360-5377.

Ruff, I., Baranyai, A., Spohr, E. and Heinzinger, K. (1989). J. Chem. Phys. **91**: 3148

Ruff, I. and Diestler, D. J. (1990). J. Chem. Phys. **93**: 2032

Russ, C., Brunner, M., Bechinger, C. and Grünberg, H. H. v. (2005). Europhys. Lett. **69**: 468.

Rutledge, G. C. (2001). Phys. Rev. E **63**: 021111

Sagui, C. and Darden, T. A. (1999). **28**: 155.

Santos, A., Yuste, S. B. and De Haro, M. L. (1999). Molecular Physics **96**(1): 1-5.

Santos, A., Yuste, S. B. and de Haro, M. L. (2002). Journal of Chemical Physics **117**(12): 5785-5793.

Schommers, W. (1973). Phys. Lett. A **43**: 157.

Shenvi, N., Geremia, J. M. and Rabitz, H. (2004). J. Chem. Phys. **120**: 9942.

Sindelka, M. and Boublik, T. (1998). Fluid Phase Equilibria **143**(1-2): 13-27.

Slattery, W. L., Doolen, G. D. and DeWitt, H. E. (1980). Phys. Rev. A **21**: 2087

Soper, A. K. (1995). Chem. Phys. **202**: 295

Soper, A. K. (2001). Mol. Phys. **99**: 1503

Soper, A. K. (2005). Phys. Rev. B **72**: 104204.

Thompson, S. E. and Parthasarathy, S. (2006). Materials Today **9**: 20-24.

Tomsic, M., Bester-Rogac, M., Jamnik, A., Kunz, W., Touraud, D., Bergmann, A. and Glatter, O. (2004). Journal of Physical Chemistry B **108**(22): 7021-7032.

Tóth, G. (2001). J. Chem. Phys. **115**: 4770

Tóth, G. (2003). J. Chem. Phys. **118**: 3949

Tóth, G. and Baranyai, A. (1997). J. Chem. Phys. **107**: 7402.

Tóth, G. and Baranyai, A. (1999). Mol. Phys. **97**: 339.

Tóth, G. and Baranyai, A. (2000). J. Mol. Liq. **85**: 3.

Tóth, G. and Baranyai, A. (2000). **165**.

Viduna, D. and Smith, W. R. (2002). Molecular Physics **100**(17): 2815-2821.

Viduna, D. and Smith, W. R. (2002). Molecular Physics **100**(17): 2903-2905.

Wilding, N. B. (2003). J. Chem. Phys. **119**: 12163.

Wolfram, S. (1999). The Mathematica Book, 4th ed. Cambridge, Wolfram Media / Cambridge University Press.

Yau, D. H. L., Chan, K. Y. and Henderson, D. (1997). Molecular Physics **91**(6): 1137-1142.

Yuste, S. B., Santos, A. and de Haro, M. L. (1998). http://www.unex.es/eweb/fisteor/santos/filesRFA.html (accessed 11 02 2005).

Yuste, S. B., Santos, A. and de Haro, M. L. (1998). Journal of Chemical Physics **108**(9): 3683-3693.

Zwicker, J. and Lovett, R. (1990). J. Chem. Phys. **93**: 6752.

List of Tables

Table 2.1 Intervals for the randomly generated systems. Minima and maxima of the reduced quantities in the selected systems. See text for the reduced units. 24

Table 2.2 Merit functions of the networks trained best. The minima and maxima correspond to the best and worst reproduced pair potentials. 31

Table 3.1 System sizes of the performed simulations. The systems were printed in italics, if we were not satisfied with the achieved statistical confidence of the pair correlation functions according to the limited computation time or possible glassy phases. 39

Table 3.2 Overlap integrals to quantify the limit of the one-component hard-sphere approximation for two cases: a) $\eta_l=\eta$ of the one-component reference b) $\eta_l+\eta_s=\eta$ of the one-component reference 43

Table 3.3 Overlap integrals to quantify the theoretical methods. O is calculated between the simulated and RFA pair correlation functions of the same systems. 47

Table 3.4 Contact values of the pair correlation functions obtained by the simulation. 50

Table 4.1 Comparison of potential energy for the polynomial fits, rational function fits, tabulated methods and cubic harmonic fits. For statistical and other details see the text. 63

Table 4.2 Comparison of forces for the polynomial fits, rational function fits, tabulated methods and cubic harmonic fits. For statistical and other details see the text. 66

Table 4.3 Number of instructions for the calculation of one interaction. The unit of the operations is one multiplication or addition. 69

List of Figures

Figure 1.1 Typical Lennard-Jones (6, 12) potential 10
Figure 1.2 Diffraction, Bragg-angles..... 13
Figure 2.1 Schematic representation of an artificial neural network..... 22
Figure 2.2 Five pair potentials reproduced best for the training (a) and test (b) sets. lines: original data, markers: output data 28
Figure 2.3 Five pair potentials reproduced worst for the training (a) and test (b) sets. lines: original data, markers: output data..... 29
Figure 2.4 Distributions of the merit function for the training to all type of potentials simultaneously. 30
Figure 2.5 Five Lennard-Jones pair potentials reproduced worst from the 90% best established potentials in the training and test sets. The neural network was trained to Lennard-Jones data. a) training set, b) test set, lines: original data, markers: output data..... 32
Figure 3.1 Comparison of large-large partial pair correlation functions for the systems of η_l=0.3 and η_s=0-0.3 at size ratio of 1:5. The legend shows η_s. 42
Figure 3.2 Comparison of large-large partial pair correlation functions for the systems of $\eta_l+\eta_s$=0.5 at size ratio of 1:5. The first and second columns in the legend show η_l and η_s, respectively. 42
Figure 3.3 Comparison of the pair correlation functions obtained with simulation and theoretical methods for the system of η_l=0.3 and η_s=0.3 at size ratio of 1:5. 44
Figure 3.4 Comparison of large-large partial pair correlation functions obtained by simulations and the RFA method for the systems of η_l=0.3 and η_s=0-0.3 at size ratio of 1:5. The legend shows η_s..... 46
Figure 3.5 Partial pair correlation functions for two simulations at size ratio of 1:5. 49
Figure 4.1 Schematic representation of a periodic system. The central cell and a possible choice of an *N* cell subsystem are denoted by thick borders. 53
Figure 4.2 Pair correlation functions of Lennard-Jones systems to check the partition choice of potential energy. The system consisted of 40 Lennard-Jones particles close to the triple point. r^* is in reduced unit (Allen and Tildesley 1987). 58

Figure 4.3 Number of operations with respect to the merit function of the fits (second column of Table 4.1). Energy calculations: □ polynomials, ■ rational functions. Energy and force calculations: + polynomials, × rational functions. The data are connected with solid and dashed lines to clarify trends..... 70

Appendix

Selected parameterised functions are detailed here. The forces can be calculated as the corresponding partial derivatives of the functions. The format of the functions helps to paste them directly into C or Fortran program codes.

P-7 polynomial

$E^{\mathrm{corr}}(x, y, z) =$

```
0.00006935528245746797+z*(0.002160008742798257
5+z*(0.010077826502256547+z*(0.08211063320794686+z*(-3.4732828079283466+z*(-9.314273611252304+(14.16681370479069-
3.111726172826123*z)*z)))))+y*(0.001575589085250862+z*(-0.04946302738211095+z*(-
0.5461572397611817+z*(9.645235058499194+z*(-
12.100970067330422+z*(10.468290488143378+6.309005153979316*z)))))+y*(-
0.12396063682007008+y*(0.982231868071471+z*(6.028068019846833+z*(-17.789466989160736+(-
96.5199335183439-42.32265420041536*z)*z))+y*(-0.9283843203267544+y*(-
27.812214458566334+y*(21.30693743338091+4.520106137117244*y+5.422380955735728*z)+z*(-
33.42332128349194+10.254096587054391*z))+z*(-
9.124340828920644+z*(98.53115658291732+30.21428274012793*z))))+z*(0.469739555622663+z*(9.73056725
774623+z*(-19.326699784455506+z*(-11.37420631739499+59.2737824517607*z))))))+x*(-
0.0010769741023264508+z*(-0.03331222484387889+z*(0.12239075346337186+z*(-
8.464141594766453+z*(32.13297640506034+(5.295849771329786-
26.680059631432933*z)*z))))+y*(0.06432376081037171+z*(0.48142207166094036+z*(-
3.456674703281202+z*(-29.470389357353614+(63.421756702461884-
141.98270528757035*z)*z)))+y*(0.5714542131297252+z*(-
21.095935501594877+z*(60.85631110259151+(363.4053322178775-3.385620913504718*z)*z))+y*(-
19.141586520349748+y*(67.1392836634568+(151.95009370062058-
296.24604740512046*z)*z+y*(77.75706782621464-81.30021189598561*y+50.407253193861436*z))+z*(-
22.584361540660517+z*(-94.37317073592516+109.45214658704292*z)))))+x*(-
0.006065872558772311+z*(0.30113981770468445+z*(13.312076580467963+z*(29.617911396911396+z*(-
116.72457001390151+107.43007317129914*z))))+y*(-0.6255143663443122+z*(10.524061781990161+z*(-
8.57249448295761+z*(-
273.58407474601285+77.40904240322111*z)))+y*(21.83680155975584+z*(123.72400023547327+(-
333.2887060666331-582.325447006672*z)*z)+y*(21.326945569638394+y*(-
334.6661821372041+80.44298584353629*y-249.48921174203804*z)+z*(-
131.7347948910907+542.7920598545852*z))))+x*(0.047081255188474065+z*(-4.043834137787724+z*(-
21.28091606308152+(66.20110806516355-11.71291472658606*z)*z))+x*(-
2.634940113127114+z*(23.224190523041237+(-9.4717876144749-
184.09034470257453*z)*z)+y*(26.71884823132592+(187.66033464664753-
385.5182685340148*z)*z+y*(54.07626820719544-384.66964141880385*y+7.403361674992371*z))+x*(-
4.328124199512774+y*(-65.56910963194828+86.47299097623085*y-
119.6071622023683*z)+x*(11.549646843718113-
13.071896023932819*x+48.37625921900341*y+37.23947113930151*z)+z*(-
```

50.97079346331456+109.92078183137382*z)))+y*(-2.0733112680607295+y*(-65.23997175235587+z*(-171.0130695529695+125.31168108618317*z)+y*(222.63085930387916+250.28203045061449*y+298.4931606713601*z))+z*(-86.15037384165008+z*(230.49331958022708+575.3871498824374*z))))))

R-6/1 polynomial

$E^{\mathrm{corr}}(x,y,z)$ =(-0.0000960512168521368+z*(-0.00560704850766644 8+z*(0.0906226345763506+z*(-0.21639277687019992+z*(-3.277365176264637+(2.004882272381631-0.6512401928726239*z)*z))))+y*(-0.007829400157423769+y*(0.13798121954712936+z*(-0.8126272788330875+z*(11.060280733203957+(4.441358500803813-2.375135450021934*z)*z))+y*(-0.008639986515872387+z*(0.4009488000594831+(4.270529310803303-19.335723163875056*z)*z)+y*(-3.419015489617705+y*(-4.0272766940210065+0.8304169575975706*y-2.095211263815416*z)+z*(6.624728386093039+1.242354506552586*z))))+z*(0.1495830902516656+z*(-0.8900349273909299+z*(-1.3091761203010592+z*(-0.4933802467002245+1.8779096673018691*z)))))+x*(0.004196154599397339+z*(-0.00701069724990378+z*(-0.6943869185678819+z*(4.669182014814679+z*(3.633879338225599+1.1393052601669498*z))))+y*(-0.003933806714890548+z*(-1.0983974530865155+z*(7.788585776402839+(-2.7875144215781593+0.15939476651879092*z)*z))+y*(-1.7678580265053072+y*(1.5481102263919486+y*(13.37660017591304+13.060615130380311*y-7.923680438753291*z)+z*(-17.891239477025035+15.87717897146121*z))+z*(6.009037922746548+z*(-41.59808673171331+21.395376236167262*z))))+x*(-0.033272445690500634+z*(0.584020424351269+z*(6.523442379770122+(-12.05263249562082-4.290313716049*z)*z))+x*(-0.007809226559581139+x*(-1.580463496237746+x*(0.894523744996066+2.3599451952315995*x-12.47838559940495*y-4.2078185053570625*z)+(6.090355759865185-6.459937570553608*z)*z+y*(13.375904119024488+12.785165253609035*y+4.8327615813011 74*z))+z*(-3.0230947417126575+z*(-2.5362219185889225+6.2035397639485 87*z))+y*(-6.374217731434383+y*(-23.717360459500625+13.23809107886893*y-12.898330486490876*z)+z*(-3.7235104862966772+17.110057606860188*z)))+y*(1.02638198826738 04+z*(2.169543574904084+(-11.3101826530877+0.5010755864028 72*z)*z)+y*(14.238540869774098+z*(-1.8873012512884682+4.2782905574141 04*z)+y*(-1.3843737175442605-34.26326140949819*y+29.92194171904781*z))))))/(0.7785801748034736-1.1780056799421226*x+0.47467769284219097*y-0.03459234405920862*z)

R-4/3 polynomial

$E^{\mathrm{corr}}(x,y,z)$ =(0.276603502858369+z*(-4.561654391176089+z*(-16.848531578909338+(-496.30931123018576-330.4895352528282*z)*z))+x*(-10.889480971076306+y*(-127.82716466691214+y*(1417.2490046012413+610.8986719429023*y-250.0357830465296*z)+(-613.048742546201-2514.4374426553577*z)*z)+z*(-1.8223169952808964+z*(807.2550480845954+1081.9222907760686*z))+x*(127.33552277440984+x*(-691.0342370897528+531.4267614254303*x+130.68301042391113*y-538.1623155961747*z)+z*(229.83556002084737+39.590231961430064*z)+y*(303.4943899078969-1570.0668499481665*y+1175.7251203585736*z)))+y*(9.597851468039439+z*(56.881627373512124+z*(433.15875573482634+704.4361965120456*z))+y*(-42.95108114953812+y*(-567.0114168090131-62.94770861897754*y+52.151196418409896*z)+z*(143.00363221889194+1626.282611553713*z))))/(504.8345

7101003654+z*(-246.65352293075762+(70.6749845912373-156.47732101930012*z)*z)+y*(-256.846382918661+z*(658.3031284197873+180.52973086841993*z)+y*(298.8238930832602-489.4813036276625*y+269.0487436607461*z))+x*(-1223.2416233280578+y*(120.44506592295635+514.6341617498665*y-1534.3392088908931*z)+x*(1172.607406289951-518.2452442689786*x+124.348779121524 52*y-664.3639329999492*z)+z*(953.9993392115671+442.61872952630995*z)))

R-3/3 polynomial

$E^{\mathrm{corr}}(x, y, z)$ =(-5.6468521708440065+z*(-0.6455850054332947+(-230.05350161010912-54.17650850083276*z)*z)+y*(7.992951149397931+y*(-87.49761949669715-1949.205115702377*y+254.24445550679738*z)+z*(181.25043636088205+1407.8975806293345*z))+x*(82.85111615402805+y*(-130.87697422086134+3887.4574763651417*y-752.741678587582*z)+x*(-295.7807664461502-621.19046643253*x+375.94415623137803*y+232.01135342400596*z)+z*(-66.94866662350915+2827.9582789815113*z)))/(2144.1865380721265+y*(-3105.804366519403+y*(-378.9440833358306-1362.783553384834*y-806.3630159591904*z)+(-1028.4360620641837-1779.4015301222191*z)*z)+z*(526.7280804765501+(1974.1832548064363-296.2714329249807*z)*z)+x*(-5177.023448684519+x*(4330.109316156619-848.1190906080768*x-9626.709942481264*y-897.9320682272385*z)+z*(-607.5401961600213+558.5610651538227*z)+y*(10484.299117231063+3700.5309423965664*y+2417.052553088499*z))

P-3 polynomial

$E^{\mathrm{corr}}(x, y, z)$ =-0.002662936871531699+z*(-0.0389581124980426 2+(0.10436444596789068-2.4819620219519227*z)*z)+y*(0.1216783487028532+z*(0.8951353654415654+1.5726746317550928*z)+y*(0.6405532694957459-3.158266234200411*y+1.8729798816149168*z)) +x*(-0.028001857078216615+y*(-2.668489424206896+4.37373697972929*y-4.873877687390871*z)+x*(1.107379412223926-3.965579695813902*x+ 5.927493669743001*y+3.231343548491743*z)+z*(-0.9380907165155405+4.647533829714995*z))