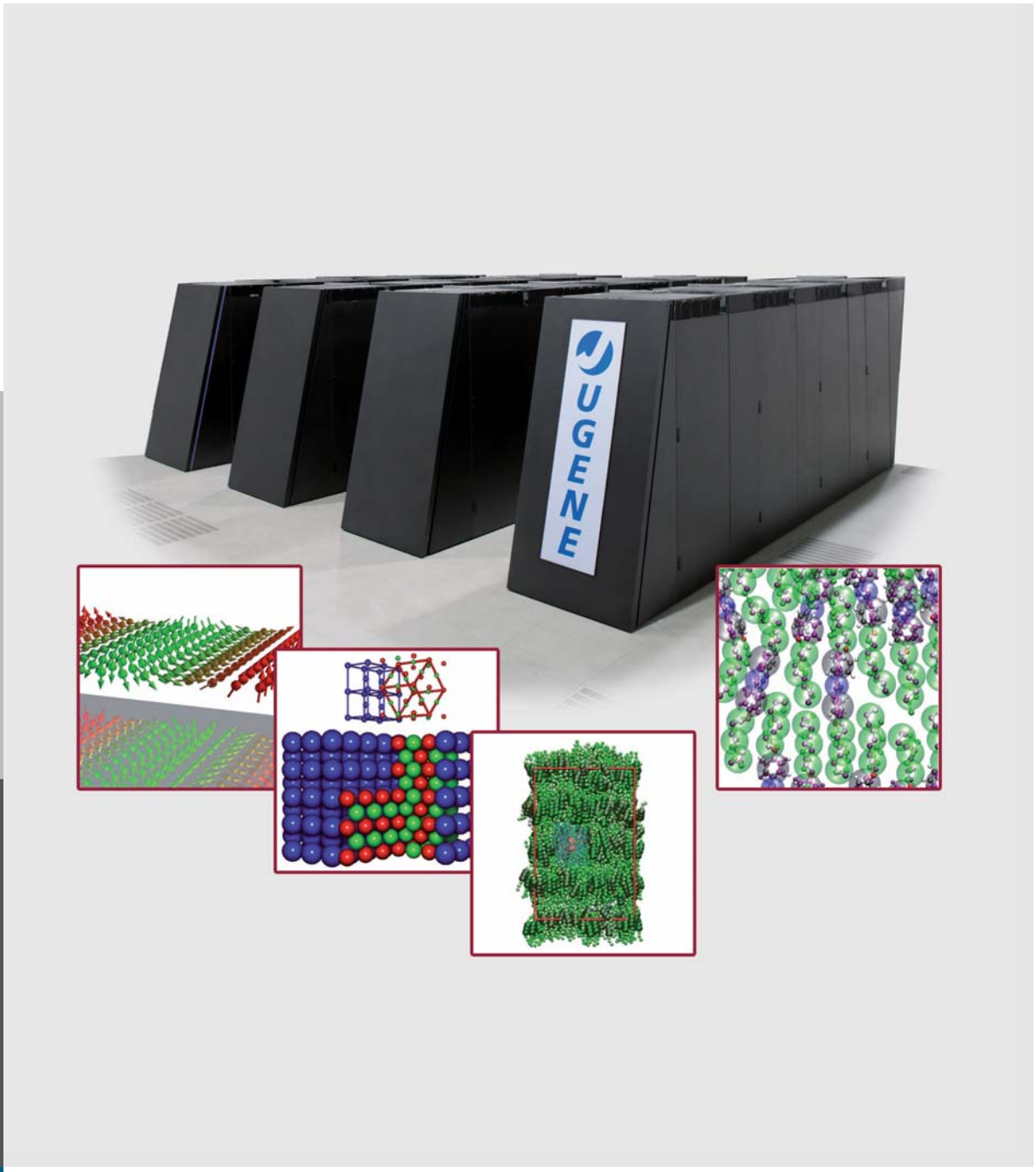


Multiscale Simulation Methods in Molecular Sciences
Lecture Notes

edited by Johannes Grotendorst, Norbert Attig, Stefan Blügel, Dominik Marx

Institute for Advanced Simulation
Jülich Supercomputing Centre

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
FORSCHUNGSZENTRUM

Publication Series of the John von Neumann Institute for Computing (NIC)
NIC Series Volume 42

Institute for Advanced Simulation (IAS)

Multiscale Simulation Methods in Molecular Sciences

edited by
Johannes Grotendorst
Norbert Attig
Stefan Blügel
Dominik Marx

Winter School, 2 - 6 March 2009
Forschungszentrum Jülich, Germany
Lecture Notes

organized by
Forschungszentrum Jülich
Ruhr-Universität Bochum

NIC Series Volume 42

ISBN 978-3-9810843-8-2

Die Deutsche Bibliothek – CIP-Cataloguing-in-Publication-Data
A catalogue record for this publication is available from Die Deutsche Bibliothek.

Publisher: Jülich Supercomputing Centre

Technical Editor: Monika Marx

Distributor: Jülich Supercomputing Centre
Forschungszentrum Jülich
52425 Jülich
Germany

Internet: www.fz-juelich.de/nic

Printer: Graphische Betriebe, Forschungszentrum Jülich

NIC Series Volume 42
ISBN 978-3-9810843-8-2

Preface

Computational techniques in the realm of molecular sciences, covering much of those parts of physics, chemistry and biology that deal with molecules, are well established in terms of extremely powerful but highly specialized approaches such as band structure calculation, quantum chemistry and biomolecular simulation. This is, for instance, nicely demonstrated by the series of Winter Schools devoted over the years to several well-defined and mature areas such as ”Quantum Chemistry” (2000), ”Quantum Many-Body Systems” (2002),”Soft Matter” (2004), and ”Nanoscience” (2006).

However, more and more problems are being tackled in experiments which have become truly molecular in the sense of accessible length and time scales using scanning probe techniques and femtosecond spectroscopy, to name but two prominent examples, which require several aspects to be covered at the same time. In most cases, it is various length scales and/or time scales covered by separate computational techniques that need to be intimately connected, or at least traversed, in order to establish a fruitful crosslink between research in the real laboratory and in the ”virtual lab”. This is precisely what the Winter School aims to address in 2009 after having covered the state of the art in many specialized areas as documented by the publication of several detailed volumes of lecture notes in the NIC publication series (available free of charge for download at www.fz-juelich.de/nic-series/).

In short, the Winter School 2009 deals with what we would like to call ”eclecticism in simulation”. The definition of eclecticism currently found in Wikipedia, is ”a conceptual approach that does not hold rigidly to a single paradigm or set of assumptions, but instead draws upon multiple theories, styles, or ideas to gain complementary insights into a subject, or applies different theories in particular cases” although not (yet) with reference to the natural sciences but only to, for example, architecture, music and psychology, perfectly describes the situation we encounter.

In particular, three topic areas will be covered focusing on how to deal with hard matter, soft matter, and bio matter where it is necessary to cope with disparate length and time scales. Aspects like coarse graining of molecular systems and solids, quantum/classical hybrid methods, embedding and multiple time step techniques, creating reactive potentials, multiscale magnetism, adaptive resolution ideas or hydrodynamic interactions will be discussed in detail. In addition, another series of lectures will be devoted to the genuine mathematical and the generic algorithmic aspects of multiscale approaches and their parallel implementation on large, multiprocessor platforms including techniques such as multigrid and wavelet transformations. Although this is beyond what can be achieved in a very systematic fashion given the breadth of the topic, introductions will be given to fundamental techniques such as molecular dynamics, Monte Carlo simulation, and electronic structure (total energy) calculations in the flavour of both wavefunction-based and density-based methods.

It is clear to the organizers that multiscale simulation is a rapidly evolving and multifaceted field that is far from being coherent and from becoming mature in the near future given the unresolved challenges of connecting, in a conceptually sound and theoretically clear-cut fashion, various length and time scales. Still, we think that the time has come to organize a Winter School on this topic in order to provide at least a glimpse of what is going on to the upcoming generation of scientists.

The scientific programme was drawn up by Johannes Grotendorst, Norbert Attig and Stefan Blügel (Forschungszentrum Jülich) and Dominik Marx (Ruhr-Universität Bochum).

The school's target audience is once again young scientists, especially PhD students and young postdocs. Because of the limited resources for the computer labs the number of participants is restricted to about 50. Applicants for the school were selected on the basis of scientific background and excellence. In spite of these restrictions, we received a wide national and international response, in most cases together with the submission of a poster abstract. This reflects the attractiveness of the programme and demonstrates the expectations of the participants that they will be able to play an active role in this high-level scientific school. We are sure that the school is stimulating for both sides, for students as well as lecturers.

Many individuals and institutions have made a significant contribution to the success of the school. First of all, we are very grateful to the lecturers for preparing the extended lecture notes in good time, in spite of the heavy work load they all have to bear. Without their efforts such an excellent reference book on multiscale simulation methods would not have been possible.

We would like to thank Forschungszentrum Jülich for financial support. We are greatly indebted to the school's secretaries Eva Mohr (handling of applications) and Erika Wittig (registration and accommodation). Special thanks go to Monika Marx for her work in compiling all the contributions and creating a high quality book from them.

Jülich and Bochum
March 2009

Johannes Grotendorst
Norbert Attig
Stefan Blügel
Dominik Marx

Contents

Methodological Foundations

Multiscale Simulation Methods for Solids and Materials

Multiscale Simulation Methods for Soft Matter and Biological Systems

Numerical Methods and Parallel Computing

Molecular Dynamics - Extending the Scale from Microscopic to Mesoscopic

Godehard Sutmann

Institute for Advanced Simulation (IAS)
Jülich Supercomputing Centre (JSC)
Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: g.sutmann@fz-juelich.de

An introduction to classical molecular dynamics simulation is presented. In addition to some historical notes, an overview is given over particle models, integrators and different ensemble techniques. In the end, methods are presented for parallelisation of short range interaction potentials. The efficiency and scalability of the algorithms on massively parallel computers is discussed with an extended version of Amdahl's law.

1 Introduction

Computer simulation methods have become a powerful tool to solve many-body problems in statistical physics[1], physical chemistry[2] and biophysics[3]. Although both the theoretical description of complex systems in the framework of statistical physics as well as the experimental techniques for detailed microscopic information are rather well developed it is often only possible to study specific aspects of those systems in great detail via simulation. On the other hand, simulations need specific input parameters that characterize the system in question, and which come either from theoretical considerations or are provided by experimental data. Having characterized a physical system in terms of model parameters, simulations are often used both to solve theoretical models beyond certain approximations and to provide a hint to experimentalists for further investigations. In the case of big experimental facilities it is often even required to prove the potential outcome of an experiment by computer simulations. In this sense it can be stated that the field of computer simulations has developed into a very important branch of science, which on the one hand helps theorists and experimentalists to go beyond their *inherent limitations* and on the other hand is a scientific field on its own. Therefore, simulation science has often been called the *third pillar* of science, complementing theory and experiment.

The traditional simulation methods for many-body systems can be divided into two classes, i.e. stochastic and deterministic simulations, which are largely represented by the Monte Carlo (MC) method[1,4] and the molecular dynamics[5,6] (MD) method, respectively. Monte Carlo simulations probe the configuration space by trial moves of particles. Within the so-called Metropolis algorithm, the energy change from step n to $n+1$ is used as a trigger to accept or reject a new configuration. Paths towards lower energy are always accepted, those to higher energy are accepted with a probability governed by Boltzmann statistics. This algorithm ensures the correct limiting distribution and properties of a given system can be calculated by averaging over all Monte Carlo moves within a given statistical ensemble (where one move means that every degree of freedom is probed once on average). In contrast, MD methods are governed by the system Hamiltonian and consequently

Hamilton's equations of motion[7,8]

$$\dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i} \qquad , \qquad \dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i} \tag{1}$$

are integrated to move particles to new positions and to assign new velocities at these new positions. This is an advantage of MD simulations with respect to MC, since not only the configuration space is probed but the whole phase space which gives additional information about the dynamics of the system. Both methods are complementary in nature but they lead to the same averages of static quantities, given that the system under consideration is ergodic and the same statistical ensemble is used.

In order to characterise a given system and to simulate its complex behavior, a model for interactions between system constituents is required. This model has to be tested against experimental results, i.e. it should reproduce or approximate experimental findings like distribution functions or phase diagrams, and theoretical constraints, i.e. it should obey certain fundamental or limiting laws like energy or momentum conservation.

Concerning MD simulations the ingredients for a program are basically threefold:
(i) As already mentioned, a model for the interaction between system constituents (atoms, molecules, surfaces etc.) is needed. Often, it is assumed that particles interact only pairwise, which is exact e.g. for particles with fixed partial charges. This assumption greatly reduces the computational effort and the work to implement the model into the program.
(ii) An integrator is needed, which propagates particle positions and velocities from time t to $t + \delta t$. It is a finite difference scheme which propagates trajectories discretely in time. The time step δt has properly to be chosen to guarantee stability of the integrator, i.e. there should be no drift in the system's energy.
(iii) A statistical ensemble has to be chosen, where thermodynamic quantities like pressure, temperature or the number of particles are controlled. The natural choice of an ensemble in MD simulations is the microcanonical ensemble (NVE), since the system's Hamiltonian without external potentials is a conserved quantity. Nevertheless, there are extensions to the Hamiltonian which also allow to simulate different statistical ensembles.

These steps essentially form the essential framework an MD simulation. Having this tool at hand, it is possible to obtain *exact* results within numerical precision. Results are only correct with respect to the model which enters into the simulation and they have to be tested against theoretical predictions and experimental findings. If the simulation results differ from *real system* properties or if they are incompatible with *solid* theoretical manifestations, the model has to be refined. This procedure can be understood as an adaptive refinement which leads in the end to an approximation of a model of the *real world* at least for certain properties. The model itself may be constructed from plausible considerations, where parameters are chosen from neutron diffraction or NMR measurements. It may also result from first principle *ab initio* calculations. Although the electronic distribution of the particles is calculated very accurately, this type of model building contains also some approximations, since many-body interactions are mostly neglected (this would increase the parameter space in the model calculation enormously). However, it often provides a good starting point for a realistic model.

An important issue of simulation studies is the accessible time- and length-scale which can be covered by microscopic simulations. Fig.1 shows a schematic representation for different types of simulations. It is clear that the more detailed a simulation technique

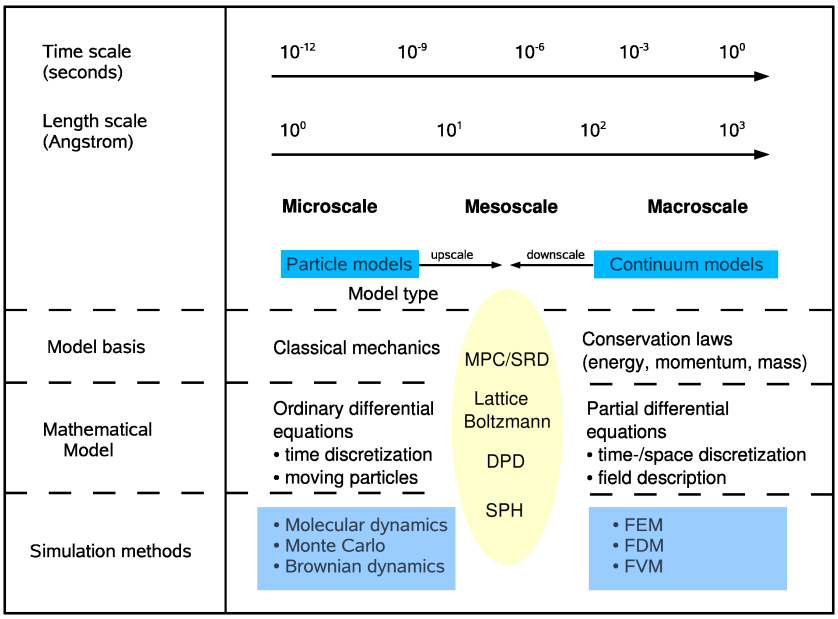


Figure 1. Schematic of different time- and length-scales, occurring from microscopic to macroscopic dimensions. Due to recent developments of techniques like Stochastic Rotation Dynamics (SRD) or Lattice Boltzmann techniques, which are designed to simulate the mesoscopic scales, there is the potential to combine different methods in a multiscale approach to cover a broad spectrum of times and lengths.

operates, the smaller is the accessibility of long times and large length scales. Therefore quantum simulations, where electronic fluctuations are taken into account, are located in the part of the diagram of very short time and length scales which are typically of the order of $\mathring{A}$ and ps. Classical molecular dynamics approximates electronic distributions in a rather coarse-grained fashion by putting either fixed partial charges on interaction sites or by adding an approximate model for polarization effects. In both cases, the time scale of the system is not dominated by the motion of electrons, but the time of intermolecular collision events, rotational motions or intramolecular vibrations, which are orders of magnitude slower than those of electron motions. Consequently, the time step of integration is larger and trajectory lengths are of order ns and accessible lengths of order $10 - 100\ \mathring{A}$. If one considers tracer particles in a solvent medium, where one is not interested in a detailed description of the solvent, one can apply Brownian dynamics, where the effect of the solvent is hidden in average quantities. Since collision times between tracer particles is very long, one may apply larger timesteps. Furthermore, since the solvent is not simulated explicitly, the lengthscales may be increased considerably. Finally, if one is interested not in a microscopic picture of the simulated system but in macroscopic quantities, the concepts of hydrodynamics may be applied, where the system properties are hidden in effective numbers, e.g. density, viscosity or sound velocity.

It is clear that the performance of particle simulations strongly depends on the computer facilities at hand. The first studies using MD simulation techniques were performed in 1957

by B. J. Alder and T. E. Wainright[9] who simulated the phase transition of a system of hard spheres. The general method, however, was presented only two years later[10]. In these early simulations, which were run on an IBM-704, up to 500 particles could be simulated, for which 500 collisions per hour were calculated. Taking into account 200000 collisions for a production run, these simulations lasted for more than two weeks. Since the propagation of hard spheres in a simulation is event driven, i.e. it is determined by the collision times between two particles, the propagation is not based on an integration of the equations of motion, but rather the calculation of the time of the next collision, which results in a variable time step in the calculations.

The first MD simulation which was applied to atoms interacting via a continuous potential was performed by A. Rahman in 1964. In this case, a model system for Argon was simulated and not only binary collisions were taken into account but the interactions were modeled by a Lennard-Jones potential and the equations of motion were integrated with a finite difference scheme. This work may be considered as seminal for dynamical calculations. It was the first work where a numerical method was used to calculate dynamical quantities like autocorrelation functions and transport coefficients like the diffusion coefficient for a realistic system. In addition, more involved characteristic functions like the dynamic van Hove function and non-Gaussian corrections to diffusion were evaluated. The calculations were performed for 864 particles on a CDC 3600, where the propagation of all particles for one time step took $\approx 45\ s$. The calculation of 50000 timesteps then took more than three weeks! [a]

With the development of faster and bigger massively parallel architectures the accessible time and length scales are increasing for all-atom simulations. In the case of classical MD simulations it is a kind of competition to break new world records by carrying out demonstration runs of larger and larger particle systems[11–14]. In a recent publication, it was reported by Germann and Kadau[15] that a trillion-atom (10^{12} particles!) simulation was run on an IBM BlueGene/L machine at Lawrence Livermore National Laboratory with 212992 PowerPC 440 processors with a total of 72 TB memory. This run was performed with the memory optimised program SPaSM[16,17] (Scalable Parallel Short-range Molecular dynamics) which, in single-precision mode, only used 44 Bytes/particle. With these conditions a simulation of a Lennard-Jones system of $N = (10000)^3$ was simulated for 40 time steps, where each time step used $\approx 50secs$ wall clock time.

Concerning the accessible time scales of all-atom simulations, a numerical study, carried out by Y. Duan and P. A. Kollman in 1998 still may be considered as a milestone in simulation science. In this work the protein folding process of the subdomain HP-36 from the villin headpiece[18,19] was simulated up to $1\ \mu s$. The protein was modelled with a 596 interaction site model dissolved in a system of 3000 water molecules. Using a timestep of integration of $2 \times 10^{-15}s$, the program was run for 5×10^8 steps. In order to perform this type of calculation, it was necessary to run the program several months on a CRAY T3D and CRAY T3E with 256 processors. It is clear that such kind of a simulation is exceptional due to the large amount of computer resources needed, but it was nonetheless a kind of milestone pointing to future simulation practices, which are nowadays still not standard, but nevertheless exceptionally applied[20].

Classical molecular dynamics methods are nowadays applied to a huge class of prob-

[a] On a standard PC this calculation may be done within less than one hour nowadays!

lems, e.g. properties of liquids, defects in solids, fracture, surface properties, friction, molecular clusters, polyelectrolytes and biomolecules. Due to the large area of applicability, simulation codes for molecular dynamics were developed by many groups. On the internet homepage of the Collaborative Computational Project No.5 (CCP5)[21] a number of computer codes are assembled for condensed phase dynamics. During the last years several programs were designed for parallel computers. Among them, which are partly available free of charge, are, e.g., Amber/Sander[22], CHARMM[23], NAMD[24], NWCHEM[25], GROMACS[26] and LAMMPS[27].

Although, with the development of massively parallel architectures and highly scalable molecular dynamics codes, it has become feasible to extend the time and length scales to *relatively* large scales, a lot of processes are still beyond technical capabilities. In addition, the time and effort for running these simulations is enormous and it is certainly still far beyond of standard. A way out of this dilemma is the invention of new simulation of methodological approaches. A method which has attracted a lot of interest recently is to coarse grain all-atom simulations and to approximate interactions between individual atoms by interactions between whole groups of atoms, which leads to a smaller number of degrees of freedom and at the same time to a smoother energy surface, which on the one hand side increases the computation between particle interactions and on the other hand side allows for a larger time step, which opens the way for simulations on larger time and length scales of physical processes[28]. Using this approach, time scales of more than 1 $\mu secs$ can now be accessed in a fast way[29,30], although it has to be pointed out that coarse grained force fields have a very much more limited range of application than all-atom force fields. In principle, the coarse graining procedure has to be outlined for every different thermodynamic state point, which is to be considered in a simulation and from that point of view coarse grain potentials are not transferable in a straight forward way as it is the case for a wide range of all-atom force field parameters.

2 Models for Particle Interactions

A system is completely determined through it's Hamiltonian $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$, where $\mathcal{H}_0$ is the *internal* part of the Hamiltonian, given as

$$\mathcal{H}_0 = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + \sum_{i<j}^{N} u(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i<j}^{N} u^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \ldots \tag{2}$$

where $\mathbf{p}$ is the momentum, m the mass of the particles and u and $u^{(3)}$ are pair and three-body interaction potentials. $\mathcal{H}_1$ is an external part, which can include time dependent effects or external sources for a force. All simulated objects are defined within a model description. Often a precise knowledge of the interaction between atoms, molecules or surfaces are not known and the model is constructed in order to describe the main features of some observables. Besides boundary conditions, which are imposed, it is the model which completely determines the system from the physical point of view. In classical simulations the *objects* are most often described by point-like centers which interact through pair- or multibody interaction potentials. In that way the highly complex description of electron dynamics is abandoned and an effective picture is adopted where the main features like the hard core of a particle, electric multipoles or internal degrees of freedom of a molecules are

modeled by a set of parameters and (most often) analytical functions which depend on the mutual position of particles in the configuration. Since the parameters and functions give a complete information of the system's energy as well as the force acting on each particle through $\mathbf{F} = -\nabla U$, the combination of parameters and functions is also called a *force field*[31]. Different types of force field were developed during the last ten years. Among them are e.g. MM3[32], MM4[33], Dreiding[34], SHARP[35], VALBON[36], UFF[37], CFF95[38], AMBER[39] CHARMM[40], OPLS[41] and MMFF[42].

There are major differences to be noticed for the potential forms. The first distinction is to be made between pair- and multibody potentials. In systems with no constraints, the interaction is most often described by pair potentials, which is simple to implement into a program. In the case where multibody potentials come into play, the counting of interaction partners becomes increasingly more complex and dramatically slows down the execution of the program. Only for the case where interaction partners are known in advance, e.g. in the case of torsional or bending motions of a molecule can the interaction be calculated efficiently by using neighbor lists or by an intelligent way of indexing the molecular sites.

A second important difference between interactions is the spatial extent of the potential, classifying it into short and long range interactions. If the potential drops down to zero faster than r^{-d}, where r is the separation between two particles and d the dimension of the problem, it is called short ranged, otherwise it is long ranged. This becomes clear by considering the integral

$$I = \int \frac{dr^d}{r^n} = \begin{cases} \infty & : n \leq d \\ \text{finite} & : n > d \end{cases} \tag{3}$$

i.e. a particles' potential energy gets contributions from *all particles of the universe* if $n \leq d$, otherwise the interaction is bound to a certain region, which is often modeled by a spherical interaction range. The long range nature of the interaction becomes most important for potentials which only have potential parameters of the same sign, like the gravitational potential where no screening can occur. For Coulomb energies, where positive and negative charges may compensate each other, long range effects may be of minor importance in some systems like molten salts.

There may be different terms contributing to the interaction potential between particles, i.e. there is no universal expression, as one can imagine for first principles calculations. In fact, contributions to interactions depend on the model which is used and this is the result of collecting various contributions into different terms, coarse graining interactions or imposing constraints, to name a few. Generally one can distinguish between bonded and non-bonded terms, or intra- and inter-molecular terms. The first class denotes all contributions originating between particles which are closely related to each other by constraints or potentials which guaranty defined particles as close neighbors. The second class denotes interactions between particles which can *freely* move, i.e. there are no defined neighbors, but interactions simply depend on distances.

A typical form for a (so called) force field (e.g. AMBER[22]) looks as follows

$$\begin{aligned} \mathcal{U} = & \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] \\ & + \sum_{i<j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right] + \sum_{\text{H}-\text{bonds}} \left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right] + \sum_{i<j} \frac{q_i q_j}{r_{ij}} \end{aligned} \tag{4}$$

In the following, short- and long-range interaction potentials and methods are briefly described in order to show differences in their algorithmical treatment.

In the following two examples shall illustrate the different treatment of short- and long range interactions.

2.1 Short Range Interactions

Short range interactions offer the possibility to take into account only neighbored particles up to a certain distance for the calculation of interactions. In that way a cutoff radius is introduced beyond of which mutual interactions between particles are neglected. As an approximation one may introduce *long range corrections* to the potential in order to compensate for the neglect of explicit calculations. The whole short range potential may then be written as

$$U = \sum_{i<j}^{N} u(r_{ij}|r_{ij} < R_c) + U_{lrc} \tag{5}$$

The long-range correction is thereby given as

$$U_{lrc} = 2\pi N \rho_0 \int_{R_c}^{\infty} dr\ r^2 g(r) u(r) \tag{6}$$

where ρ_0 is the number density of the particles in the system and $g(r) = \rho(r)/\rho_0$ is the radial distribution function. For computational reasons, $g(r)$ is most often only calculated up to R_c, so that in practice it is assumed that $g(r) = 1$ for $r > R_c$, which makes it possible for many types of potentials to calculate U_{lrc} analytically.

Besides internal degrees of freedom of molecules, which may be modeled with short range interaction potentials, it is first of all the excluded volume of a particle which is of importance. A finite diameter of a particle may be represented by a steep repulsive potential acting at short distances. This is either described by an exponential function or an algebraic form, $\propto r^{-n}$, where $n \geq 9$. Another source of short range interaction is the van der Waals interaction. For neutral particles these are the London forces arising from induced dipole interactions. Fluctuations of the electron distribution of a particle give rise to fluctuating dipole moments, which on average compensate to zero. But the instantaneous created dipoles induce also dipoles on neighbored particles which attract each other $\propto r^{-6}$. Two common forms of the resulting interactions are the Buckingham potential

$$u^B_{\alpha\beta}(r_{ij}) = A_{\alpha\beta} e^{-B_{\alpha\beta} r_{ij}} - \frac{D_{\alpha\beta}}{r^6_{ij}} \tag{7}$$

and the Lennard-Jones potential

$$u^{LJ}_{\alpha\beta}(r_{ij}) = 4\epsilon_{\alpha\beta} \left(\left(\frac{\sigma_{\alpha\beta}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{ij}} \right)^{6} \right) \tag{8}$$

which are compared in Fig.2. In Eqs.7,8 the indices α, β indicate the species of the particles, i.e. there are parameters A, B, D in Eq.7 and ϵ, σ in Eq.8 for intra-species interactions ($\alpha = \beta$) and cross species interactions ($\alpha \neq \beta$). For the Lennard-Jones potential the parameters have a simple physical interpretation: ϵ is the minimum potential energy,

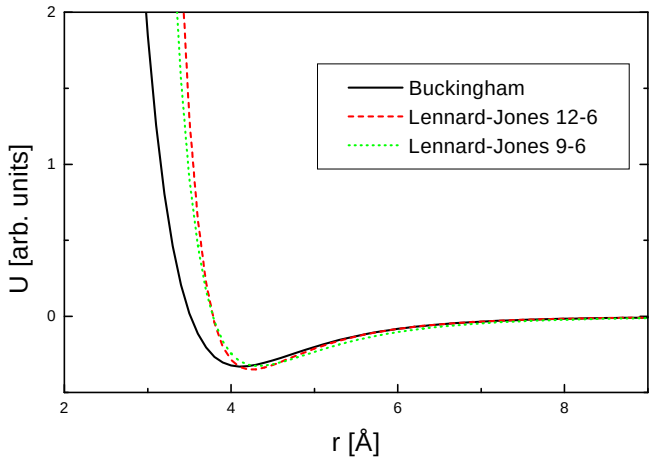


Figure 2. Comparison between a Buckingham-, Lennard-Jones (12-6) and Lennard-Jones (9-6) potential.

located at $r = 2^{1/6}\sigma$ and σ is the diameter of the particle, since for $r < \sigma$ the potential becomes repulsive. Often the Lennard-Jones potential gives a reasonable approximation of a *true* potential. However, from exact quantum ab initio calculations an exponential type repulsive potential is often more appropriate. Especially for dense systems the too steep repulsive part often leeds to an overestimation of the pressure in the system. Since computationally the Lennard-Jones interaction is quite attractive the repulsive part is sometimes replaced by a weaker repulsive term, like $\propto r^{-9}$. The Lennard-Jones potential has another advantage over the Buckingham potential, since there are combining rules for the parameters. A common choice are the Lorentz-Berelot combining rules

$$\sigma_{\alpha\beta} = \frac{\sigma_{\alpha\alpha} + \sigma_{\beta\beta}}{2} \quad , \quad \epsilon_{\alpha\beta} = \sqrt{\epsilon_{\alpha\alpha}\epsilon_{\beta\beta}} \tag{9}$$

This combining rule is, however, known to overestimate the well depth parameter. Two other commonly known combining rules try to correct this effect, which are Kong[43] rules

$$\sigma_{\alpha\beta} = \left[\frac{1}{2^{13}} \frac{\epsilon_{\alpha\alpha}\sigma_{\alpha\alpha}^{12}}{\sqrt{\epsilon_{\alpha\alpha}\sigma_{\alpha\alpha}^{6}\epsilon_{\beta\beta}\sigma_{\beta\beta}^{6}}} \left(1 + \left(\frac{\epsilon_{\beta\beta}\sigma_{\beta\beta}^{12}}{\epsilon_{\alpha\alpha}\sigma_{\alpha\alpha}^{12}} \right)^{1/13} \right)^{13} \right]^{1/6} \tag{10}$$

$$\epsilon_{\alpha\beta} = \frac{\sqrt{\epsilon_{\alpha\alpha}\sigma_{\alpha\alpha}^{6}\epsilon_{\beta\beta}\sigma_{\beta\beta}^{6}}}{\sigma_{\alpha\beta}^{6}} \tag{11}$$

and the Waldman-Kagler[44] rule

$$\sigma_{\alpha\beta} = \left(\frac{\sigma_{\alpha\alpha}^{6} + \sigma_{\beta\beta}^{6}}{2} \right)^{1/6} \quad , \quad \epsilon_{\alpha\beta} = \frac{\sqrt{\epsilon_{\alpha\alpha}\sigma_{\alpha\alpha}^{6}\epsilon_{\beta\beta}\sigma_{\beta\beta}^{6}}}{\sigma_{\alpha\beta}^{6}} \tag{12}$$

In a recent study[45] of Ar-Kr and Ar-Ne mixtures, these combining rules were tested and it was found that the Kong rules give the best agreement between simulated and experimental

pressure-density curves. An illustration of the different combining rules is shown in Fig.3 for the case of an Ar-Ne mixture.

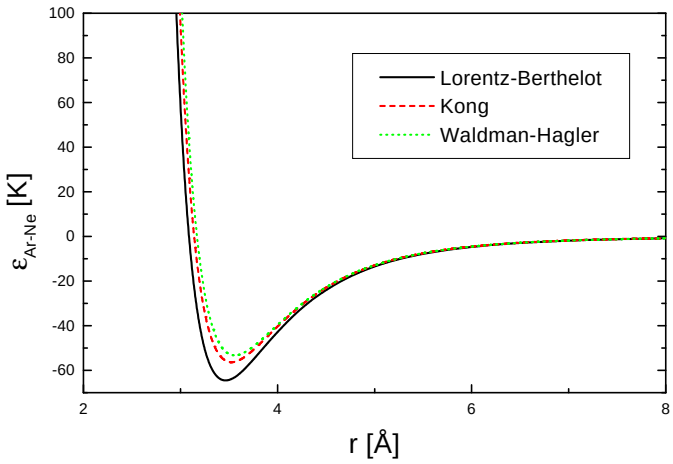


Figure 3. Resulting cross-terms of the Lennard-Jones potential for an Ar-Ne mixture. Shown is the effect of different combining rules (Eqs.9-12). Parameters used are $\sigma_{Ar} = 3.406\ \mathring{A}, \epsilon_{Ar} = 119.4\ K$ and $\sigma_{Ne} = 2.75\ \mathring{A}, \epsilon_{Ne} = 35.7\ K$.

Since there are only relatively few particles which have to be considered for the interaction with a tagged particle (i.e. those particles within the cutoff range), it would be a computational bottleneck if in any time step all particle pairs would have to be checked whether they lie inside or outside the interaction range. This becomes more and more a problem as the number of particles increases. A way to overcome this bottleneck is to introduce list techniques. The first implementation dates back to the early days of molecular dynamics simulations. In 1967, Verlet introduced a list[46], where at a given time step all particle pairs were stored within a range $R_c + R_s$, where R_s is called the skin radius and which serves as a reservoir of particles, in order not to update the list in each time step (which would make the list redundant). Therefore, in a force routine, not all particles have to tested, whether they are in a range $r_{ij} < R_c$, but only those particle pairs, stored in the list. Since particles are moving during the simulation, it is necessary to update the list from time to time. A criterion to update the list could be, e.g.

$$\max_i |\mathbf{r}_i(t) - \mathbf{r}_i(t_0)| \geq \frac{R_s}{2} \tag{13}$$

where t_0 is the time from the last list update. This ensures that particles cannot move from the outside region into the cutoff sphere without being recognized. This technique, though efficient, has still complexity $\mathcal{O}(N^2)$, since at an update step, *all* particle pairs have to be checked for their mutual distances. Another problem arises when simulating many particles, since the memory requirements are relatively large (size of the list is $4\pi(R_c + R_s)^3 \rho\, N/3$). There is, of course also the question, how large the skin radius should be chosen. Often, it is chosen as $R_s = 1.5\sigma$. In Ref.[47] it was shown that an optimal choice

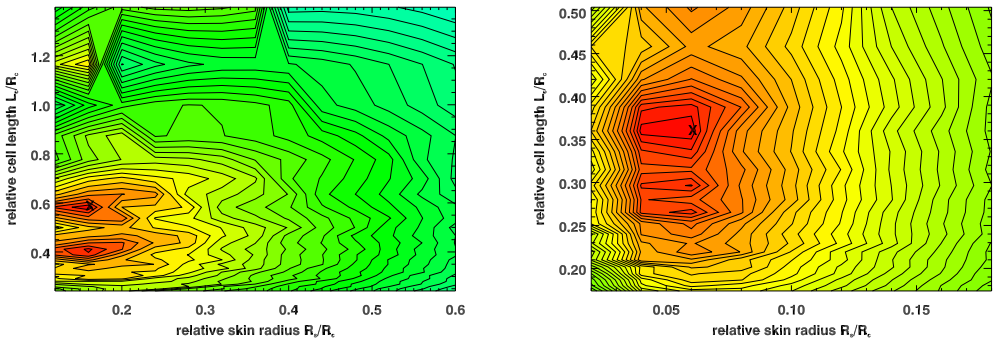


Figure 4. Contour plots of the performance for the combination of linked-cell and Verlet list as a function of the cell length and the size of the skin radius. Crosses mark the positions predicted from an optimization procedure[48]. Test systems were composed of 4000 Lennard-Jones particles with $R_c = 2.5\,\sigma$ at temperature $T = 1.4\,\epsilon/k_B$. Left: $\rho = 0.75/\sigma^3$. Right: $\rho = 2.0/\sigma^3$.

strongly depends on the number of particles in the system and an optimization procedure was outlined.

An alternative list technique, which scales linearly with the number of particles is the linked-cell method[49,50]. The linked-cell method starts with subdividing the whole system into cubic cells and sorting all particles into these cells according to their position. The size of the cells, L_c, is chosen to be $L_c \leq L_{Box}/\text{floor}(L_{Box}/R_c)$, where L_{Box} is the length of the simulation box. All particles are then sorted into a list array of length N. The list is organized in a way that particles, belonging to the same cell are linked together, i.e. the entry in the list referring to a particle points directly to the entry of a next particle inside the same cell. A zero entry in the list stops the search in the cell and a next cell is checked for entries. This technique not only has computational complexity of $O(N)$, since the sorting into the cells and into the N-dimensional array is of $O(N)$, but also has memory requirements which only grow linearly with the number of particles. These features make this technique very appealing. However, the technique is not well vectorizable and also the addressing of next neighbors in the cells require indirect access (e.g. `i=index(i)`), which may lead to cache misses. In order not to miss any particle pair in the interactions every box has to have a neighbor region in each direction which extends to R_c. In the case, where $L_c \geq R_c$, every cell is surrounded by 26 neighbor cells in three dimensional systems. This gives rise to the fact that the method gives only efficiency gains if $L_{Box} \geq 4R_c$, i.e. subdividing each box direction into more than 3 cells. In order to approximate the cutoff sphere in a better way by cubic cells, one may reduce the cell size and simultaneously increasing the total number of cells. In an optimization procedure[47], it was found that a reduction of cell sizes to $L_c = R_c/2$ or even smaller often gives very much better results.

It is, of course, possible to combine these list techniques, i.e. using the linked-cell technique in the update step of the Verlet list. This reduces the computational complexity of the Verlet list to $\mathcal{O}(N)$ while fully preserving the efficiency of the list technique. It is also possible to model the performance of this list combination and to optimize the length of the cells and the size of the skin radius. Figure 4 shows the result of a parameter study,

where the performance of the list was measured as a function of (L_c, R_s). Also shown is the prediction of parameters coming out of an optimization procedure[48].

2.2 Long Range Interactions

Long range interactions essentially require to take all particle pairs into account for a proper treatment of interactions. This may become a problem, if periodic boundary conditions are imposed to the system, i.e. formally simulating an infinite number of particles (no explicit boundaries imply infinite extend of the system). Therefore one has to devise special techniques to treat this situation. On the other hand one also has to apply fast techniques to overcome the inherent $\mathcal{O}(N^2)$ complexity of the problem, since for large numbers of particles this would imply an intractable computational bottleneck. In general one can classify algorithms for long range interactions into the following system:

- Periodic boundary conditions
  - Grid free algorithms, e.g. Ewald summation method[51–53]
  - Grid based algorithms, e.g. Smoothed Particle Mesh Ewald[54,55], Particle-Particle Particle-Mesh method[56–58]
- Open boundary conditions
  - Grid free algorithms, e.g. Fast Multipole Method[59–64] (FMM), Barnes-Hut Tree method[65,66]
  - Grid based algorithms, e.g. Particle-Particle Particle-Multigrid method[67] (P^3Mg), Particle Mesh Wavelet method[68] (PMW)

In the following two important members of these classes will be described, the Ewald summation method and the Fast Multipole Method.

2.2.1 Ewald Summation Method

The Ewald summation method originates from crystal physics, where the problem was to determine the Madelung constant[69], describing a factor for an effective electrostatic energy in a perfect periodic crystal. Considering the electrostatic energy of a system of N particles in a cubic box and imposing periodic boundary conditions, leads to an equivalent problem. At position $\mathbf{r}_i$ of particle i, the electrostatic potential, $\phi(\mathbf{r}_i)$, can be written down as a lattice sum

$$\phi(\mathbf{r}_i) = \sum_{\mathbf{n}}^{\dagger} \sum_{j=1}^{N} \frac{q_j}{\|\mathbf{r}_{ij} + \mathbf{n}L\|} \tag{14}$$

where $\mathbf{n} = (n_x, n_y, n_z), n_\alpha \in \mathbb{Z}$ is a vector along cartesian coordinates and L is the length of the simulation box. The sign "†" means that $i \neq j$ for $\|\mathbf{n}\| = 0$.

Eq. (14) is conditionally convergent, i.e. the result of the outcome depends on the order of summation. Also the sum extends over infinite number of lattice vectors, which means that one has to modify the procedure in order to get an absolute convergent sum and to get it fast converging. The original method of Ewald consisted in introducing a convergence

factor e^{-ns}, which makes the sum absolute convergent; then transforming it into different fast converging terms and then putting s in the convergence factor to zero. The final result of the calculation can be easier understood from a physical picture. If every charge in the system is screened by a counter charge of opposite sign, which is smeared out, then the potential of this composite charge distribution becomes short ranged (it is similar in electrolytic solutions, where ionic charges are screened by counter charges - the result is an exponentially decaying function, the Debye potential[70]). In order to compensate for the added charge distribution it has to be subtracted again. The far field of a localized charge distribution is, however, again a Coulomb potential. Therefore this term will be long ranged. There would be nothing gained if one would simply sum up these different terms. The efficiency gain shows up, when one calculates the short range interactions as direct particle-particle contributions in real space, while summing up the long range part of the smeared charge cloud in reciprocal Fourier space. Choosing as the smeared charge distribution a Gaussian charge cloud of half width $1/\alpha$ the corresponding expression for the energy becomes

$$\begin{aligned}\phi(\mathbf{r}_i) = &\sum_{\mathbf{n}}{}^{\dagger} \sum_{j=1}^{N} q_j \frac{\operatorname{erfc}(\alpha \|\mathbf{r}_{ij} + \mathbf{n}L\|)}{\|\mathbf{r}_{ij} + \mathbf{n}L\|} \\ &+ \frac{4\pi}{L^3} \sum_{\mathbf{k}\neq 0} \sum_{j=1}^{N} \frac{q_j}{\|\mathbf{k}\|^2} e^{-\|\mathbf{k}\|^2/4\alpha^2} e^{i\mathbf{k}\mathbf{r}_{ij}} - q_i \frac{2\alpha}{\sqrt{\pi}}\end{aligned} \tag{15}$$

The last term corresponds to a self energy contribution which has to be subtracted, as it is considered in the Fourier part. Eq. (15) is an exact equivalent of Eq. (14), with the difference that it is an absolute converging expression. Therefore nothing would be gained without further approximation. Since the complimentary error function can be approximated for large arguments by a Gaussian function and the k-space parts decreases like a Gaussian, both terms can be approximated by stopping the sums at a certain lattice vector $\mathbf{n}$ and a maximal k-value k_{max}. The choice of parameters depends on the error, $\epsilon = \exp(-p^2)$, which one accepts to tolerate. Setting the error tolerance p and choosing the width of the counter charge distribution, one gets

$$R_c^2 + \frac{\log(R_c)}{\alpha^2} = \frac{1}{\alpha^2}(p^2 - \log(2)) \tag{16}$$

$$k_{max}^2 + 8\alpha^2 \log(k_{max}) = 4\alpha^2 p^2 + \log\left(\frac{4\pi}{L^3}\right) \tag{17}$$

This can be solved iteratively or if one is only interested in an approximate estimate for the error, i.e. neglecting logarithmic terms, one gets

$$R_c = \frac{p}{\alpha} \tag{18}$$

$$k_{max} = 2\alpha p \tag{19}$$

Using this error estimate and furthermore introducing execution times, spent for the real- and reciprocal-space part, it is possible to show that parameters R_c, α and k_{max} can be chosen to get a complexity of $\mathcal{O}(N^{3/2})$ for the Ewald sum[71,72]. In this case, parameters

are

$$\frac{R_c}{L} \approx \sqrt{\frac{\pi}{N^{1/3}}} \quad , \quad \alpha L \approx \frac{L k_{max}}{2\pi} = \sqrt{\pi N^{1/3}} \tag{20}$$

Figure 5 shows the contributions of real- and reciprocal parts in Eq. (15), as a function of the spreading parameter α, where an upper limit in both the real- and reciprocal-contributions was applied. In the real-space part usually one restricts the sum to $|\mathbf{n}| = 0$ and applies a spherical cutoff radius, R_c. For fixed values of R_c and k_{max} there is a broad plateau region, where the two terms add up to a constant value. Within this plateau region, a value for α should be chosen. Often it is chosen according to $\alpha = 5/L$. Also shown is the potential energy of a particle, calculated with the Ewald sum. It is well observed that due to the periodicity of the system the potential energy surface is not radial symmetric, which may cause problems for small numbers of particles in the system.

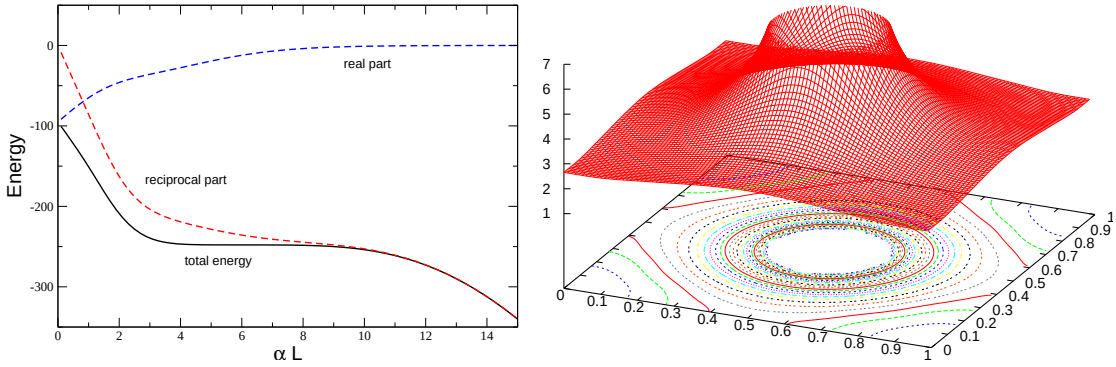


Figure 5. Left: Dependence of the calculated potential on the choice of the scaled inverse width, $\alpha\, L$, of the smeared counter charge distribution. Parameters for this test were $N = 152$, $R_c = 0.5\, L$ and $k_{max} L/2\pi = 6$. Right: Surface plot and contours for the electrostatic potential of a charge, located in the center of the simulation volume. Picture shows the xy-plane for $z = L/2$. Parameters were $R_c = 0.25\, L$, $\alpha\, L = 12.2$ and $k_{max} L/2\pi = 6$.

The present form of the Ewald sum gives an exact representation of the potential energy of point like charges in a system with periodic boundary conditions. Sometimes the charge distribution in a molecule is approximated by a point dipole or higher multipole moments. A more general form of the Ewald sum, taking into account arbitrary point multipoles was given in Ref.[73]. The case, where also electronic polarizabilities are considered is given in Ref.[74].

In certain systems, like in molten salts or electrolyte solutions, the interaction between charged species may approximated by a screened Coulomb potential, which has a Yukawa-like form

$$U = \frac{1}{2} \sum_{i,j=1}^{N} q_i q_j \frac{e^{-\kappa \|\mathbf{r}_{ij}\|}}{\|\mathbf{r}_{ij}\|} \tag{21}$$

The parameter κ is the inverse Debye length, which gives a measure of screening strength in the system. If $\kappa < 1/L$ the potential is short ranged and usual cut-off methods may

be used. Instead, if $\kappa > 1/L$, or generally if $u(r = L/2)$ is larger than the prescribed uncertainties in the energy, the minimum image convention in combination with truncation methods fails and the potential must be treated in a more rigorous way, which was proposed in Ref.[75], where an extension of the Ewald sum for such Yukawa type potentials was developed.

2.2.2 The Fast Multipole Method

In open geometries there is no lattice summation, but only the sum over all particle pairs in the whole system. The electrostatic energy at a particle's position is therefore simply calculated as

$$\phi(\mathbf{r}_i) = \sum_{j=1}^{N} \frac{q_j}{\|\mathbf{r}_i - \mathbf{r}_j\|} \tag{22}$$

Without further approximation this is always an $\mathcal{O}(N^2)$ algorithm since there are $N(N-1)/2$ interactions to consider in the system (here Newton's third law was taken into account). The idea of a multipole method is to group particles which are far away from a tagged particle together and to consider an effective interaction of a particle with this particle group[76–78]. The physical space is therefore subdivided in a hierarchical way, where the whole system is considered as level 0. Each further level is constructed by dividing the length in each direction by a factor of two. The whole system is therefore subdivided into a hierarchy of boxes where each *parent box* contains eight *children boxes*. This subdivision is performed at maximum until the level, where each particle is located in an individual box. Often it is enough to stop the subdivision already at a lower level.

In the following it is convenient to work in spherical coordinates. The main principle of the method is that the interaction between two particles, located at $\mathbf{r} = r, \theta, \varphi$ and $\mathbf{a} = (a, \alpha, \beta)$ can be written as a multipole expansion[79]

$$\frac{1}{\|\mathbf{r} - \mathbf{a}\|} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{(l-|m|)!}{(l+|m|)!} \frac{a^l}{r^{l+1}} P_{lm}(\cos\alpha)\, P_{lm}(\cos\theta)\, e^{-im(\beta-\varphi)} \tag{23}$$

where $P_{lm}(x)$ are associated Legendre polynomials[80]. This expression requires that $a/r < 1$ and this gives a lower limit for the so called *well separated* boxes. This makes it necessary to have at least one box between a tagged box and the zone, where contributions can be expanded into multipoles. Defining the operators

$$O_{lm}(\mathbf{a}) = a^l\, (l-|m|)!\, P_{lm}(\cos\alpha)\, e^{-im\beta} \tag{24}$$

$$M_{lm}(\mathbf{r}) = \frac{1}{r^{l+1}} \frac{1}{(l+|m|)!}\, P_{lm}(\cos\theta)\, e^{im\varphi} \tag{25}$$

with which Eq. (23) may simply be rewritten in a more compact way, it is possible to write further three operators, which are needed, in a compact scheme, i.e.

1.) a translation operator, which relates the multipole expansion of a point located at $\mathbf{a}$ to a multipole expansion of a point located at $\mathbf{a} + \mathbf{b}$

$$O_{lm}(\mathbf{a}+\mathbf{b}) = \sum_{j=0}^{l} \sum_{k=-l}^{j} A_{jk}^{lm}(\mathbf{b})\, O_{jk}(\mathbf{a}) \qquad , \qquad A_{jk}^{lm}(\mathbf{b}) = O_{l-j,m-k}(\mathbf{b}) \tag{26}$$

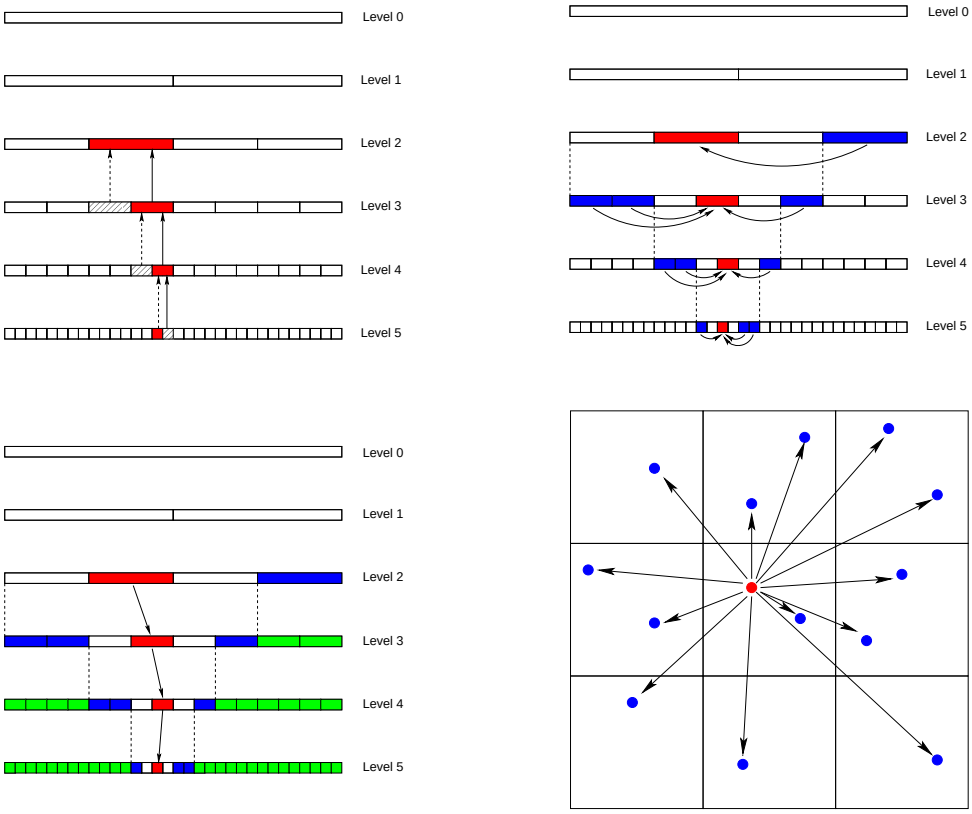


Figure 6. Schematic of different passes in the Fast Multipole Method. Upper left: Pass 1, evaluation of multipole terms in finest subdivision and translating information upwards the tree. Upper right: Pass 2, transforming multipole expansions in well separated boxes into local Taylor expansions. Lower left: Pass 3, transferring multipole expansions downwards the tree, thus collecting information of the whole system, except nearest neighbor boxes. Lower right: Pass 5, direct calculation of particle-particle interactions in local and nearest neighbor boxes.

2.) a transformation operator, which transforms a multipole expansion centered at the origin into a Taylor expansion centered at location $\mathbf{b}$

$$M_{lm}(\mathbf{a}-\mathbf{b}) = \sum_{j=0}^{l}\sum_{k=-l}^{j} B_{jk}^{lm}(\mathbf{b})\, O_{jk}(\mathbf{a}) \qquad , \qquad B_{jk}^{lm}(\mathbf{b}) = M_{l+j,m+k}(\mathbf{b}) \qquad (27)$$

3.) a translation operator, which translates a Taylor expansion of a point $\mathbf{r}$ about the origin into a Taylor expansion of $\mathbf{r}$ about a point $\mathbf{b}$

$$M_{lm}(\mathbf{r}-\mathbf{b}) = \sum_{j=0}^{l}\sum_{k=-l}^{j} C_{jk}^{lm}(\mathbf{b})\, M_{jk}(\mathbf{r}) \qquad , \qquad C_{jk}^{lm}(\mathbf{b}) = A_{jk}^{lm}(\mathbf{b}) \qquad (28)$$

The procedure to calculate interactions between particles is then subdivided into five passes. Figure 6 illustrates four of them. The first pass consists of calculating the multipole expansions in the lowest level boxes (finest subdivision). Using the translation operator $O_{lm}(\mathbf{a}+\mathbf{b})$, the multipole expansions are translated into the center of their parent boxes and summed up. This procedure is repeated then subsequently for each level, until level 2 is reached, from where no further information is passed to a coarser level. In pass 2, using operator $M_{lm}(\mathbf{a}-\mathbf{b})$, multipole expansions are translated into Taylor expansions in a box from well separated boxes, whose parent boxes are nearest neighbor boxes. Well separated means, that for all particles in a given box the multipole expansion in a separated box is valid. Since the applicability of Eq. (23) implies $r > a$, well separateness means on level l that boxes should be separated by a distance 2^{-l}. This also explains, why there is no need to transfer information higher than level 2, since from there on it is not possible to have well separated boxes anymore, i.e. multipole expansions are not valid any more. In pass 3, using the operator $M_{lm}(\mathbf{a}-\mathbf{b})$, this information is then translated downwards the tree, so that finally on the finest level all multipole information is known in order to interact individual particles with expansions, originating from all other particles in the system which are located in well separated boxes of the finest level. In pass 4 this interaction between individual particles and multipoles is performed. Finally in pass 5, explicit pair-pair interactions are calculated between particles in a lowest level box and those which are in nearest neighbor boxes, i.e. those boxes which are not called well separated.

It can be shown[61] that each of the steps performed in this algorithm is of order $\mathcal{O}(N)$, making it an optimal method. Also the error made by this method can be controlled rather reliably[64]. A very conservative error estimate is thereby given as[76,61,81]

$$\left|\phi(r) - \frac{q}{\|\mathbf{r}-\mathbf{a}\|}\right| \leq \frac{|q|}{r-a}\left(\frac{a}{r}\right)^{p+1} \tag{29}$$

At the current description the evaluation of multipole terms scales as $\mathcal{O}(l_{max}^4)$, when l_{max} is the largest value of l in the multipole expansion, Eq.(23). A faster version which scales as $\mathcal{O}(l_{max}^3)$ and therefore strongly reducing the prefactor of the overall scheme, was proposed in Ref.[62], where multipoles are evaluated in a rotated coordinate frame, which makes it possible to reduce calculations to Legendre polynomials and not requiring associated Legendre polynomials.

Also to mention is that there are approaches to extend the Fast Multipole Method to periodic systems[82,83].

2.3 Coarse Grain Methods

The force field methods mentioned so far treat molecules on the atomic level, i.e. resolving heavy atoms, in most cases also hydrogens, explicitly. In the case, where flexible molecular bonds, described e.g. by harmonic potentials, are considered the applied time step is of the order of $\delta t \approx 10^{-15}\, secs$. Considering physical phenomena like self assembling of lipid molecules[84,85], protein folding or structure formation in macromolecular systems[86–88], which take place on time scales of microseconds to seconds or even longer, the number of timesteps would exceed the current computational capacities. Although these phenomena all have an underlying microscopic background, the fast dynamics of e.g. hydrogen vibrations are not directly reflected in the overall process. This lead to the

idea to either freeze certain degrees of freedom, as it is done for e.g. rigid water models[89–92], or to take several degrees of freedom only into account effectively via a pseudo potential, which reflects the behavior of whole groups of atoms. It is the latter approach which is now known as coarse graining[28,93,94] of molecular potentials and which opens the accessibility of a larger time and length scale. Mapping groups of atoms to one pseudo atom, or interaction site, leads already to an effective increase of the specific volume of the degrees of freedom. Therefore, the same number of degrees of freedom of a coarse grain model, compared with a conventional force field model, would directly lead to larger spatial scale, due to the increase of volume of each degree of freedom. On the other hand, comparing a conventional system before and after coarse graining, the coarse grained system could cover time scales longer by a factor of 100-1000 or even longer compared with a conventional force field all-atom model (the concrete factor certainly depends on the level of coarse graining).

Methodologies for obtaining coarse grain models of a system often start from an atomistic all-atom model, which adequately describes phase diagrams or other physical properties of interest. On a next level, groups of atoms are collected and an effective non-bonded interaction potential may be obtained by calculating potential energy surfaces of these groups and to parametrize these potentials to obtain analytical descriptions. Therefore, distribution functions of small atomic groups are taken into account (at least implicitly) which in general depend on the thermodynamic state point. For bonded potentials between groups of atoms, a normal mode analysis may be performed in order to get the most important contributions to vibrational-, bending- or torsional-modes.

In principle, one is interested in reducing the number of degrees of freedom by separating the problem space into coordinates which are *important* and those which are *unimportant*. Formally, this may be expressed through a set of coordinates $\{\mathbf{r}\} \in \mathbb{R}^{n_i}$ and $\{\tilde{\mathbf{r}}\} \in \mathbb{R}^{n_u}$, where n_i and n_u are the number of degrees of important and unimportant degrees of freedom, respectively. Consequently, the system Hamiltonian may be written as $H = H(r_1, \ldots, r_{n_i}, \tilde{r}_1, \ldots, \tilde{r}_{n_i})$. From these considerations one may define a *reduced* partition function, which results from integrating out all unimportant degrees of freedom

$$Z = \int dr_1, \ldots, dr_{n_i}, d\tilde{r}_1, \ldots, d\tilde{r}_{n_i} \exp\left\{-\beta H(r_1, \ldots, r_{n_i}, \tilde{r}_1, \ldots, \tilde{r}_{n_u})\right\} \quad (30)$$

$$= \int dr_1, \ldots, dr_{n_i}, d\tilde{r}_1, \ldots, d\tilde{r}_{n_u} \exp\left\{-\beta H^{CG}(r_1, \ldots, r_{n_i})\right\} \quad (31)$$

where a coarse grain Hamiltonian has been defined

$$H^{CG}(r_1, \ldots, r_{n_i}) = -\log \int \tilde{r}_1, \ldots, d\tilde{r}_{n_i} \exp\left\{-\beta H(r_1, \ldots, r_{n_i}, \tilde{r}_1, \ldots, \tilde{r}_{n_u})\right\} \quad (32)$$

which corresponds to the potential of mean force and which is the free energy of the non-important degrees of freedom. Since the Hamiltonian describes only a subset of degrees of freedom, thermodynamic properties, derived from this Hamiltonian will be different than obtained from the full Hamiltonian description (e.g. pressure will correspond to the osmotic pressure and not to the thermodynamic pressure). This has to be taken into account when simulating in different ensembles or if experimental thermodynamic properties should be reproduced by simulation.

The coarse grained Hamiltonian is still a multi-body description of the system, which is hard to obtain numerically. Therefore, it is often approximated by a pair-potential, which

is considered to contribute the most important terms

$$H^{CG}(r_1, \ldots, r_{n_i}) \approx \sum_{i>j} V_{ij}(r_{ij}) \qquad , \qquad r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\| \tag{33}$$

According to the uniqueness theorem of Henderson[95], in a liquid where particles interact only through pair interactions, the pair distribution function $g(r)$ determines up to a constant uniquely the pair interaction potential V_{ij}. Therefore, V_{ij} may be obtained pointwise by reverting the radial pair distribution function[96–98], e.g. by reverse Monte Carlo techniques[99] or dynamic iterative refinement[100]. This approach directly confirms what was stated in Sec. 1 about the limited applicability of coarse grained potentials. It is clear that for different temperatures, pressures or densities the radial distribution functions of e.g. cation-cation, cation-anion and anion-anion distributions in electrolytic solutions will be different. If one wants to simulate ions in an effective medium (continuum solvent), the potential, which is applied in the simulation will depend on the thermodynamic state point and therefore has to be re-parametrized for every different state point.

3 The Integrator

The propagation of a classical particle system can be described by the temporal evolution of the phase space variables $(\mathbf{p}, \mathbf{q})$, where the phase space $\Gamma(\mathbf{p}, \mathbf{q}) \in \mathbb{R}^{6N}$ contains all possible combinations of momenta and coordinates of the system. The exact time evolution of the system is thereby given by a flow map

$$\Phi_{\delta t, \mathcal{H}} : \mathbb{R}^{6N} \rightarrow \mathbb{R}^{6N} \tag{34}$$

which means

$$\Phi_{\delta t, \mathcal{H}}(\mathbf{p}(t), \mathbf{q}(t)) = (\mathbf{p}(t) + \delta\mathbf{p}, \mathbf{q}(t) + \delta\mathbf{q}) \tag{35}$$

where

$$\mathbf{p} + \delta\mathbf{p} = \mathbf{p}(t + \delta t) \qquad , \qquad \mathbf{q} + \delta\mathbf{q} = \mathbf{q}(t + \delta t) \tag{36}$$

For a nonlinear many-body system, the equations of motion cannot be integrated exactly and one has to rely on numerical integration of a certain order. Propagating the coordinates by a constant step size h, a number of different finite difference schemes may be used for the integration. But there are a number of requirements, which have to be fulfilled in order to be useful for molecular dynamics simulations. An integrator, suitable for many-body simulations should fulfill the following requirements:

- Accuracy, i.e. the solution of an analytically solvable test problem should be as close as possible to the numerical one.
- Stability, i.e. very long simulation runs should produce physically relevant trajectories, which are not governed by numerical artifacts
- Conservativity, there should be no drift or divergence in conserved quantities, like energy, momentum or angular momentum

- Reversibility, i.e. it should have the same temporal structure as the underlying equations
- Effectiveness, i.e. it should allow for large time steps without entering instability and should require a minimum of force evaluations, which usually need about 95 % of CPU time per time step
- Symplecticity, i.e. the geometrical structure of the phase space should be conserved

It is obvious that the numerical flow, $\phi_{\delta t,\mathcal{H}}$, of a finite difference scheme will not be fully equivalent to $\Phi_{\delta t,\mathcal{H}}$, but the system dynamics will be described correctly if the items above will be fulfilled.

In the following the mentioned points will be discussed and a number of different integrators will be compared.

3.1 Basic Methods

The most simple integration scheme is the Euler method, which may be constructed by a first order difference approximation to the time derivative of the phase space variables

$$\mathbf{p}_{n+1} = \mathbf{p}_n - \delta t \frac{\partial}{\partial \mathbf{q}} \mathcal{H}(\mathbf{p}_n, \mathbf{q}_n) \tag{37}$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \delta t \frac{\partial}{\partial \mathbf{p}} \mathcal{H}(\mathbf{p}_n, \mathbf{q}_n) \tag{38}$$

where δt is the step size of integration. This is equivalent to a Taylor expansion which is truncated after the first derivative. Therefore, it is obvious that it is of first order. Knowing all variables at step n, this scheme has all relevant information to perform the integration. Since only information from one time step is required to do the integration, this scheme is called the one-step explicit Euler scheme. The basic scheme, Eqs. (37,38) may also be written in different forms.

The implicit Euler method

$$\mathbf{p}_{n+1} = \mathbf{p}_n - \delta t \frac{\partial}{\partial \mathbf{q}} \mathcal{H}(\mathbf{p}_{n+1}, \mathbf{q}_{n+1}) \tag{39}$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \delta t \frac{\partial}{\partial \mathbf{p}} \mathcal{H}(\mathbf{p}_{n+1}, \mathbf{q}_{n+1}) \tag{40}$$

can only be solved iteratively, since the derivative on the right-hand-side (*rhs*) is evaluated at the coordinate positions on the left-hand-side (*lhs*).

An example for a so called partitioned Runge-Kutta method is the *velocity implicit method*

$$\mathbf{p}_{n+1} = \mathbf{p}_n - \delta t \frac{\partial}{\partial \mathbf{q}} \mathcal{H}(\mathbf{p}_{n+1}, \mathbf{q}_n) \tag{41}$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \delta t \frac{\partial}{\partial \mathbf{p}} \mathcal{H}(\mathbf{p}_{n+1}, \mathbf{q}_n) \tag{42}$$

Since the Hamiltonian usually splits into kinetic $\mathcal{K}$ and potential $\mathcal{U}$ parts, which only depend on one phase space variable, i.e.

$$\mathcal{H}(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \mathbf{p}^T \, \mathbf{M}^{-1} \, \mathbf{p} + \mathcal{U}(\mathbf{q}) \tag{43}$$

where $\mathbf{M}^{-1}$ is the inverse of the diagonal mass matrix, this scheme may also be written as

$$\mathbf{p}_{n+1} = \mathbf{p}_n - \delta t \frac{\partial}{\partial \mathbf{q}} \mathcal{U}(\mathbf{q}_n) \tag{44}$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \frac{\delta t}{m} \mathbf{p}_{n+1} \tag{45}$$

showing that it is not necessary to solve it iteratively.

Obviously this may be written as a *position implicit method*

$$\mathbf{p}_{n+1} = \mathbf{p}_n - \delta t \frac{\partial}{\partial \mathbf{q}} \mathcal{U}(\mathbf{q}_{n+1}) \tag{46}$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \frac{\delta t}{m} \mathbf{p}_n \tag{47}$$

Applying first Eq. (47) and afterwards Eq. (46) also this variant does not require an iterative procedure.

All of these schemes are first order accurate but have different properties, as will be shown below. Before discussing these schemes it will be interesting to show a higher order scheme, which is also based on a Taylor expansion. First write down expansions

$$\mathbf{q}(t+\delta t) = \mathbf{q}(t) + \delta t\, \dot{\mathbf{q}}(t) + \frac{1}{2}\, \delta t^2\, \ddot{\mathbf{q}}(t) + O(\delta t^3) \tag{48}$$

$$= \mathbf{q}(t) + \frac{\delta t}{m}\, \mathbf{p}(t) + \frac{1}{2m}\, \delta t^2\, \dot{\mathbf{p}}(t) + O(\delta t^3) \tag{49}$$

$$\mathbf{p}(t+\delta t) = \mathbf{p}(t) + \delta t\, \dot{\mathbf{p}}(t) + \frac{1}{2}\, \delta t^2\, \ddot{\mathbf{p}}(t) + O(\delta t^3) \tag{50}$$

$$= \mathbf{p}(t) + \frac{\delta t}{2}\, (\dot{\mathbf{p}}(t) + \dot{\mathbf{p}}(t+\delta t)) + O(\delta t^3) \tag{51}$$

where in Eq. (49), the relation $\dot{\mathbf{q}} = \mathbf{p}/m$ was used and in Eq. (51) a first order Taylor expansion for $\dot{\mathbf{p}}$ was inserted. From these expansions a simple second order, one-step splitting scheme may be written as

$$\mathbf{p}_{n+1/2} = \mathbf{p}_n + \frac{\delta t}{2}\, \mathbf{F}(\mathbf{q}_n) \tag{52}$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \frac{\delta t}{m}\, \mathbf{p}_{n+1/2} \tag{53}$$

$$\mathbf{p}_{n+1} = \mathbf{p}_{n+1/2} + \frac{\delta t}{2}\, \mathbf{F}(\mathbf{q}_{n+1}) \tag{54}$$

where the relation $\dot{\mathbf{p}} = -\partial \mathcal{H}/\partial \mathbf{q} = \mathbf{F}$ was used. This scheme is called the *Velocity Verlet* scheme. In a pictorial way it is sometimes described as half-kick, drift, half-kick, since the first step consists in applying forces for half a time step, second step consists in free flight of a particle with momentum $\mathbf{p}_{n+1/2}$ and the last step applies again a force for half a time step. In practice, forces only need to be evaluated once in each time step. After having calculated the new positions, $\mathbf{q}_{n+1}$, forces are calculated for the last integration step. They are, however, stored to be used in the first integration step as *old* forces in the next time step of the simulation.

This algorithm comes also in another flavor, called the *Position Verlet* scheme. It can be expressed as

$$\mathbf{q}_{n+1/2} = \mathbf{q}_n + \frac{\delta t}{2m}\,\mathbf{p}_n \tag{55}$$

$$\mathbf{p}_{n+1} = \mathbf{p}_n + \delta t\,\mathbf{F}(\mathbf{q}_{n+1/2}) \tag{56}$$

$$\mathbf{q}_{n+1/2} = \mathbf{q}_{n+1/2} + \frac{\delta t}{2m}\,\mathbf{p}_{n+1} \tag{57}$$

In analogy to the description above this is sometimes described as half-drift, kick, half-drift. Using the relation $\mathbf{p} = \dot{\mathbf{q}}/m$ and expressing this as a first order expansion, it is obvious that $\mathbf{F}(\mathbf{q}_{n+1/2}) = \mathbf{F}((\mathbf{q}_n + \mathbf{q}_{n+1})/2)$ which corresponds to an implicit midpoint rule.

3.2 Operator Splitting Methods

A more rigorous derivation, which in addition leads to the possibility of splitting the propagator of the phase space trajectory into several time scales, is based on the phase space description of a classical system. The time evolution of a point in the $6N$ dimensional phase space is given by the Liouville equation

$$\mathbf{\Gamma}(t) = e^{i\mathcal{L}t}\,\mathbf{\Gamma}(0) \tag{58}$$

where $\mathbf{\Gamma} = (\mathbf{q}, \mathbf{p})$ is the $6N$ dimensional vector of generalized coordinates, $\mathbf{q} = \mathbf{q}_1, \ldots, \mathbf{q}_N$, and momenta, $\mathbf{p} = \mathbf{p}_1, \ldots, \mathbf{p}_N$. The Liouville operator, $\mathcal{L}$, is defined as

$$i\mathcal{L} = \{\ldots, \mathcal{H}\} = \sum_{j=1}^{N}\left(\frac{\partial \mathbf{q}_j}{\partial t}\frac{\partial}{\partial \mathbf{q}_j} + \frac{\partial \mathbf{p}_j}{\partial t}\frac{\partial}{\partial \mathbf{p}_j}\right) \tag{59}$$

In order to construct a discrete timestep integrator, the Liouville operator is split into two parts, $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$, and a Trotter expansion[101] is performed

$$e^{i\mathcal{L}\delta t} = e^{i(\mathcal{L}_1+\mathcal{L}_2)\delta t} \tag{60}$$

$$= e^{i\mathcal{L}_1\delta t/2}e^{i\mathcal{L}_2\delta t}e^{i\mathcal{L}_1\delta t/2} + \mathcal{O}(\delta t^3) \tag{61}$$

The partial operators can be chosen to act only on positions and momenta. Assuming usual cartesian coordinates for a system of N free particles, this can be written as

$$i\mathcal{L}_1 = \sum_{j=1}^{N}\mathbf{F}_j\frac{\partial}{\partial \mathbf{p}_j} \tag{62}$$

$$i\mathcal{L}_2 = \sum_{j=1}^{N}\mathbf{v}_j\frac{\partial}{\partial \mathbf{r}_j} \tag{63}$$

Applying Eq.60 to the phase space vector $\mathbf{\Gamma}$ and using the property $e^{a\partial/\partial x}f(x) = f(x+a)$ for any function f, where a is independent of x, gives

$$\mathbf{v}_i(t+\delta t/2) = \mathbf{v}(t) + \frac{\mathbf{F}_i(t)}{m_i}\frac{\delta t}{2} \tag{64}$$

$$\mathbf{r}_i(t+\delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t+\delta t/2)\delta t \tag{65}$$

$$\mathbf{v}_i(t+\delta t) = \mathbf{v}_i(t+\delta t/2) + \frac{\mathbf{F}_i(t+\delta t)}{m_i}\frac{\delta t}{2} \tag{66}$$

which is the velocity Verlet algorithm, Eqs. 52-54. In the same spirit, another algorithm may be derived by simply changing the definitions for $\mathcal{L}_1 \rightarrow \mathcal{L}_2$ and $\mathcal{L}_2 \rightarrow \mathcal{L}_1$. This gives the so called *position Verlet algorithm*

$$\mathbf{r}_i(t + \delta t/2) = \mathbf{r}_i(t) + \mathbf{v}(t)\frac{\delta t}{2} \tag{67}$$

$$\mathbf{v}_i(t + \delta t) = \mathbf{v}(t) + \frac{\mathbf{F}_i(t + \delta t/2)}{m_i} \tag{68}$$

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t + \delta t/2) + (\mathbf{v}(t) + \mathbf{v}_i(t + \delta t))\frac{\delta t}{2} \tag{69}$$

Here the forces are calculated at intermediate positions $\mathbf{r}_i(t+\delta t/2)$. The equations of both the velocity Verlet and the position Verlet algorithms have the property of propagating velocities or positions on half time steps. Since both schemes decouple into an applied force term and a *free flight* term, the three steps are often called *half-kick*/*drift*/*half kick* for the velocity Verlet and correspondingly *half-drift*/*kick*/*half-drift* for the position Verlet algorithm.

Both algorithms, the velocity and the position Verlet method, are examples for symplectic algorithms, which are characterized by a volume conservation in phase space. This is equivalent to the fact that the Jacobian matrix of a transform $x' = f(x, p)$ and $p' = g(x, p)$ satisfies

$$\begin{pmatrix} f_x & f_p \\ g_x & g_p \end{pmatrix} \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} f_x & f_p \\ g_x & g_p \end{pmatrix} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \tag{70}$$

Any method which is based on the splitting of the Hamiltonian, is symplectic. This does not yet, however, guarantee that the method is also time reversible, which may be also be considered as a strong requirement for the integrator. This property is guaranteed by symmetric methods, which also provide a better numerical stability[102]. Methods, which try to enhance the accuracy by taking into account the particles' history (multi-step methods) tend to be incompatible with symplecticness[103, 104], which makes symplectic schemes attractive from the point of view of data storage requirements. Another strong argument for symplectic schemes is the so called *backward error analysis*[105–107]. This means that the trajectory produced by a discrete integration scheme, may be expressed as the solution of a perturbed ordinary differential equation whose *rhs* can formally be expressed as a power series in δt. It could be shown that the system, described by the ordinary differential equation is Hamiltonian, if the integrator is symplectic[108, 109]. In general, the power series in δt diverges. However, if the series is truncated, the trajectory will differ only as $\mathcal{O}(\delta t^p)$ of the trajectory, generated by the symplectic integrator on timescales $\mathcal{O}(1/\delta t)$[110].

3.3 Multiple Time Step Methods

It was already mentioned that the rigorous approach of the decomposition of the Liouville operator offers the opportunity for a decomposition of time scales in the system. Supposing that there are different time scales present in the system, e.g. fast intramolecular vibrations and slow domain motions of molecules, then the factorization of Eq.60 may be written in

a more general way

$$e^{i\mathcal{L}\Delta t} = e^{i\mathcal{L}_1^{(s)}\Delta t/2} e^{i\mathcal{L}_1^{(f)}\Delta t/2} e^{i\mathcal{L}_2\delta t} e^{i\mathcal{L}_1^{(f)}\Delta t/2} e^{i\mathcal{L}_1^{(s)}\Delta t/2} \tag{71}$$

$$= e^{i\mathcal{L}_1^{(s)}\Delta t/2} \left\{ e^{i\mathcal{L}_1^{(f)}\delta t/2} e^{i\mathcal{L}_2\delta t} e^{i\mathcal{L}_1^{(f)}\delta t/2} \right\}^p e^{i\mathcal{L}_1^{(s)}\Delta t/2} \tag{72}$$

where the time increment is $\Delta t = p\delta$. The decomposition of the Liouville operator may be chosen in the convenient way

$$i\mathcal{L}_1^{(s)} = \mathbf{F}_i^{(s)} \frac{\partial}{\partial \mathbf{p}_i} \quad , \quad i\mathcal{L}_1^{(f)} = \mathbf{F}_i^{(f)} \frac{\partial}{\partial \mathbf{p}_i} \quad , \quad i\mathcal{L}_2 = \mathbf{v}_i \frac{\partial}{\partial \mathbf{q}_i} \tag{73}$$

where the superscript (s) and (f) mean slow and fast contributions to the forces. The idea behind this decomposition is simply to take into account contributions from slowly varying components only every p'th timestep with a large time interval. Therefore, the force computation may be considerably speeded up in the the $p-1$ intermediate force computation steps. In general, the scheme may be extended to account for more time scales. Examples for this may be found in Refs.[111–113]. One obvious problem, however, is to separate the timescales in a proper way. The scheme of Eq.72 is *exact* if the time scales decouple completely. This, however, is very rarely found and most often timescales are coupled due to nonlinear effects. Nevertheless, for the case where Δt is not very much larger than δt ($p \approx 10$), the separation may be often justified and lead to stable results. Another criteria for the separation is to distinguish between long range and short range contributions to the force. Since the magnitude and the fluctuation frequency is very much larger for the short range contributions this separation makes sense for speeding up computations including long range interactions[114].

The method has, however, its limitations[115,116]. As described, a particle gets every n'th timestep a *kick* due to the slow components. It was reported in literature that this may excite a system's resonance which will lead to strong artifacts or even instabilities[117,118]. Recently different schemes were proposed to overcome these resonances by keeping the property of symplecticness[119–125].

3.4 Stability

Performing simulations of stable many-body systems for long times should produce configurations which are in thermal equilibrium. This means that system properties, e.g. pressure, internal energy, temperature etc. are fluctuating around constant values. To measure these equilibrium properties it should not be relevant where to put the time origin from where configurations are considered to calculate average quantities. This requires that the integrator should propagate phase space variables in such a way that small fluctuations do not lead to a diverging behavior of a system property. This is a kind of minimal requirement in order to simulate any physical system without a domination of numerical artifacts. It is clear, however, that any integration scheme will have its own stability range depending on the step size δt. This is a kind of sampling criterion, i.e. if the step size is too large, in order to resolve details of the energy landscape, an integration scheme may end in instability.

For linear systems it is straight forward to analyze the stability range of a given numerical scheme. Consider e.g. the harmonic oscillator, for which the equations of motion may be written as $\dot{q}(t) = p(t)$ and $\dot{p}(t) = -\omega^2 q(t)$, where ω is the vibrational frequency and it

is assumed that it oscillates around the origin. The exact solution of this problem may be written as

$$\begin{pmatrix} \omega\, q(t) \\ p(t) \end{pmatrix} = \begin{pmatrix} \cos\omega t & \sin\omega t \\ -\sin\omega t & \cos\omega t \end{pmatrix} \begin{pmatrix} \omega\, q(0) \\ p(0) \end{pmatrix} \tag{74}$$

For a numerical integrator the stepwise solution may be written as

$$\begin{pmatrix} \omega\, q_{n+1} \\ p_{n+1} \end{pmatrix} = \mathbf{M}(\delta t) \begin{pmatrix} \omega\, q_n \\ p_n \end{pmatrix} \tag{75}$$

where $\mathbf{M}(\delta t)$ is a propagator matrix. It is obvious that any stable numerical scheme requires eigenvalues $|\lambda(\mathbf{M})| \leq 1$. For $|\lambda| > 1$ the scheme will be unstable and divergent, for $|\lambda| < 1$ it will be stable but will exhibit friction, i.e. will loose energy. Therefore, in view of the conservativity of the scheme, it will be required that $|\lambda(\mathbf{M})| = 1$.

As an example the propagator matrices for the Implicit Euler (IE) and Position Verlet (PV) algorithms are calculated as

$$\mathbf{M}_{IE}(\delta t) = \frac{1}{1+\omega^2\delta t^2} \begin{pmatrix} 1 & \omega\delta t \\ -\omega\delta t & 1 \end{pmatrix} \tag{76}$$

$$\mathbf{M}_{PV}(\delta t) = \begin{pmatrix} 1 - \frac{1}{2}\omega^2\delta t^2 & \omega\delta t\left(1 - \frac{1}{4}\omega^2\delta t^2\right) \\ -\omega\delta t & 1 - \frac{1}{2}\omega^2\delta t^2 \end{pmatrix} \tag{77}$$

It is then straight forward to calculate the eigenvalues as roots of the characteristic polynomials. The eigenvalues are then calculated as

$$\lambda_{EE} = 1 \pm i\omega\delta t \tag{78}$$

$$\lambda_{IE} = \frac{1}{1+\omega^2\delta t^2}(1 \pm i\omega\delta t) \tag{79}$$

$$\lambda_{PV} = \lambda_{VV} = \lambda_{VIE} = \lambda_{PIE} = 1 - \frac{1}{2}\omega^2\delta t^2 \left(1 \pm \sqrt{1 - \frac{4}{\omega^2\delta t^2}}\right) \tag{80}$$

This shows that the absolute values for the Explicit Euler (EE) and the Implicit Euler methods never equals one for $\delta t \neq 0$, i.e. both methods do not produce stable trajectories. This is different for the Position Verlet, the Velocity Verlet (VV), the Position Implicit Euler (PIE) and the Velocity Implicit Euler (VIE), which all have the same eigenvalues. It is found that the range of stability for all of them is in the range $\omega^2\delta t^2 < 2$. For larger values of δt the absolute values of the eigenvalues bifurcates, getting larger and smaller values than one. In Figure 7 the absolute values are shown for all methods and in in Figure 8 the imaginary versus real parts of λ are shown. For EE it is clear that the imaginary part diverges linearly with increase of δt. The eigenvalues of the stable methods are located on a circle until $\omega^2\delta t^2 = 2$. From there one branch diverges to $-\infty$, while the other decreases to zero.

As a numerical example the phase space trajectories of the harmonic oscillator for $\omega = 1$ are shown for the different methods in Figure 9. For the stable methods, results

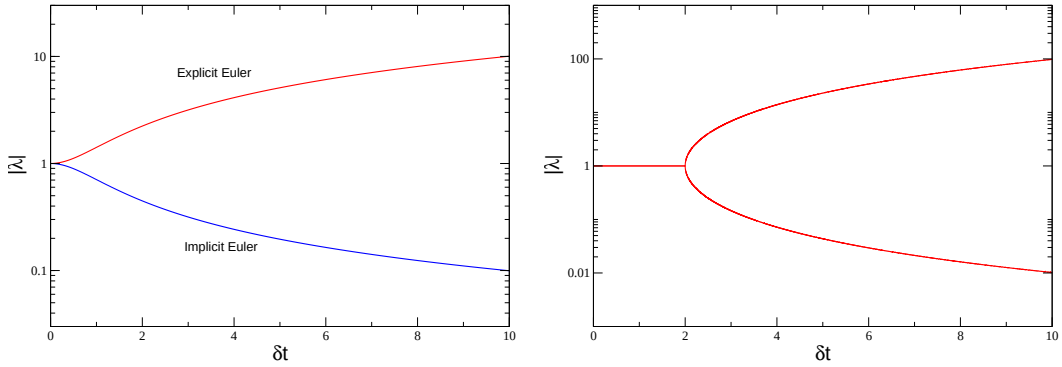


Figure 7. Absolute value of the eigenvalues λ as function of the time step δt. Left: Explicit and implicit Euler method. Right: Velocity and Position Verlet as well as Velocity Implicit and Position implicit Euler method. All methods have the eigenvalues.

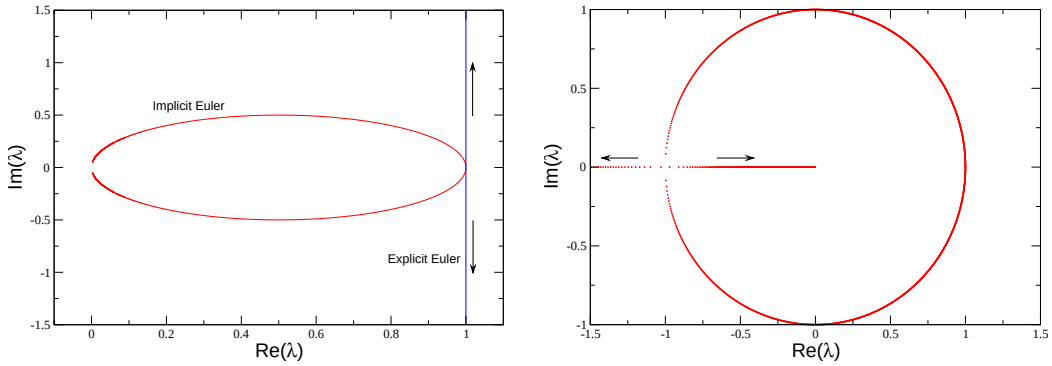


Figure 8. Imaginary versus real part of eigenvalues λ of the propagator matrices. Left: Implicit and Explicit Euler. Right: Velocity and Position Verlet as well as Velocity Implicit and Position implicit Euler method.

for a time step close to instability is shown. All different methods produce closed, stable orbits, but it is seen on the other hand that they strongly deviate from the exact solution, which is shown for reference. This demonstrates that stability is a necessary, but only a weak criterion for correct results. Numerically correct results are only obtained for much smaller time steps in the range of $\delta t \approx 0.01$. Also shown are the results for EE and IE. Here a very much smaller time step, $\delta t = 0.01$ is chosen. It is seen that the phase space trajectory of EE spirals out while the one of IE spirals in with time, showing the instable or evanescent character of the methods.

Another issue related to stability is the effect of a trajectory perturbation. If initial conditions are slightly perturbed, will a good integrator keep this trajectory close to the reference trajectory? The answer is No and it is even found that the result is not that strong dependent on the integrator. Even for integrators of high order, trajectories will not stay close to each other. The time evolution of the disturbance may be studied similar to the system trajectory. Consider the time evolution for $\Gamma + \delta\Gamma$, where $\Gamma = (\mathbf{p}, \mathbf{q})$ and

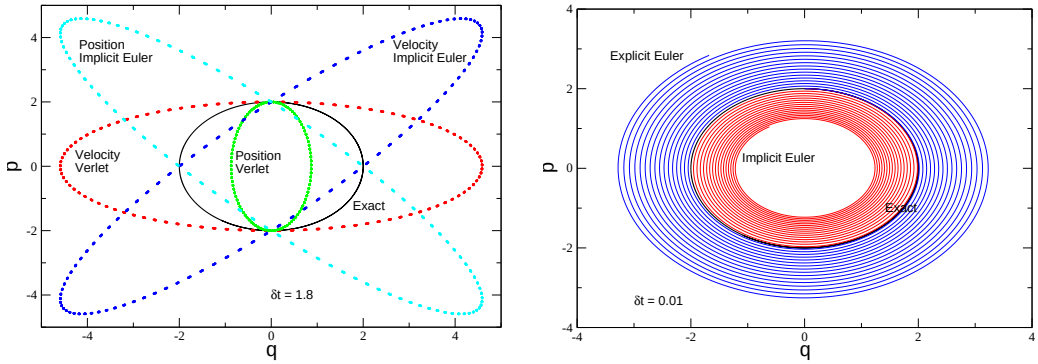


Figure 9. Phase space trajectories for the one-dimensional harmonic oscillator, integrated with the Velocity Implicit Euler, Position Implicit Euler, Velocity Verlet, Position Verlet and integration step size of $\delta t = 1.8$ (left) and the Implicit Euler and Explicit Euler and step size $\delta t = 0.01$ (right).

$\delta\Gamma = (\delta\mathbf{p}, \delta\mathbf{q})$ is a small disturbance. Then

$$\frac{d\Gamma}{dt} = \nabla_\Gamma \mathcal{H}(\Gamma) \tag{81}$$

Similarly one can write for small $\delta\Gamma$

$$\frac{d}{dt}(\Gamma + \delta\Gamma) = \nabla_\Gamma \mathcal{H}(\Gamma + \delta\Gamma) \tag{82}$$

$$= \nabla_\Gamma \mathcal{H}(\Gamma) + \nabla_\Gamma(\nabla_\Gamma \mathcal{H}(\Gamma))\delta\Gamma \tag{83}$$

where the second line is a truncated Taylor series. Comparing terms one simply gets as equation of motion for a perturbation

$$\frac{d\delta\Gamma}{dt} = \nabla_\Gamma^2 \mathcal{H}(\Gamma)\delta\Gamma \tag{84}$$

It is found that the disturbance develops exponentially, with a characteristic, system dependent exponent, which is the Ljapunov exponent[126, 127].

Now consider the following situation where identical starting configurations are taken for two simulations. They will be carried out by different yet exact algorithms, therefore leading formally to the same result. Nevertheless it may happen that different orders of floating-point operations are used in both algorithms. Due to round off errors, floating-point arithmetic is not necessarily associative, i.e. in general

$$a \mathbin{\hat{\circ}} (b \mathbin{\hat{\circ}} c) \neq (a \mathbin{\hat{\circ}} b) \mathbin{\hat{\circ}} c \tag{85}$$

where $\hat{\circ}$ is a floating-point machine operation $(+, -, /, *)$. Therefore, both simulations will be different by round off errors. According to the above discussion, this may be considered as the slightest disturbance of a system trajectory, $\delta\Gamma_{\min}$, and the question is, what effect such a round off error will have. A different method to study difference in system trajectories is the calculation of the difference

$$\gamma_x(t) = \frac{1}{3N} \sum_{i=1}^{N} \sum_{\alpha=x,y,z} (x(t) - \tilde{x}(t))^2 \tag{86}$$

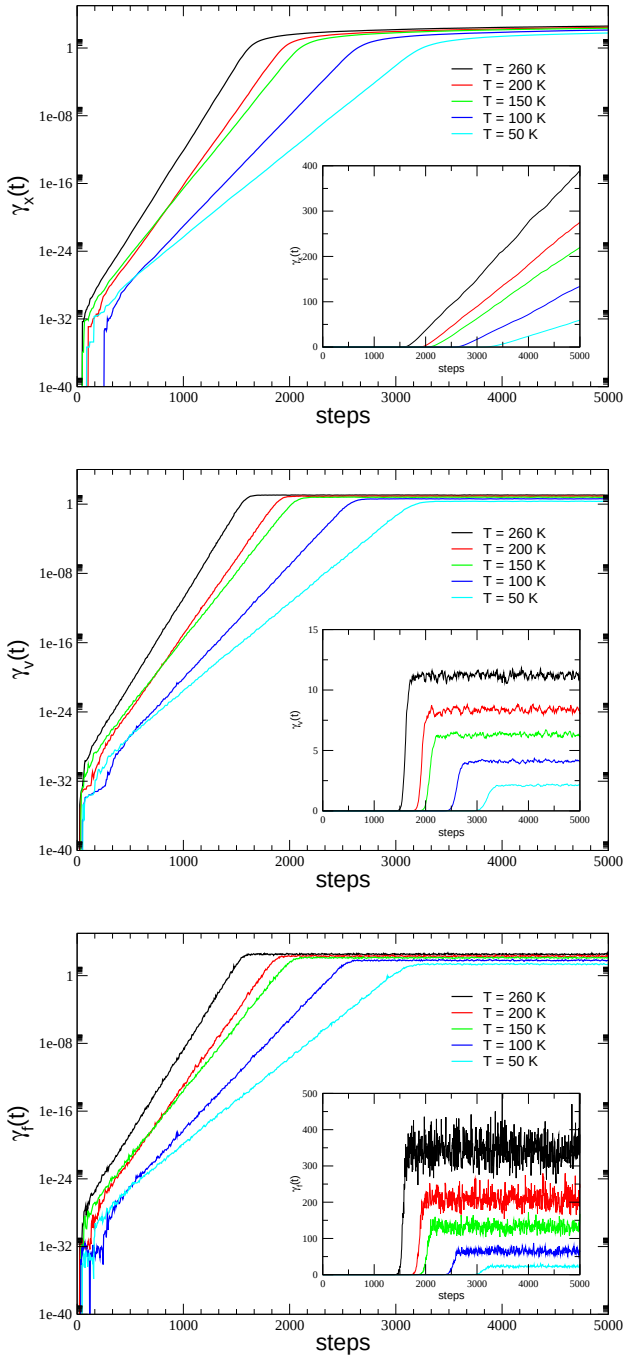


Figure 10. Divergent behavior of trajectories due to round off errors, induced by different summation order in the force routine. From top to bottom: coordinates, velocities, forces. The insets show on a linear scale the long time behavior of the trajectory differences, i.e. when the two systems get uncorrelated.

where N is the number of particles, $x(t)$ a certain property, e.g. the coordinates or momenta, and $\tilde{x}$ the same property of a disturbed trajectory. In Figure 10 results are shown for a system of Lennard-Jones particles, where the disturbance was induced by reversing the order of summation in the force routine, thereby provoking round off errors in the first time step. Shown are results for the coordinates, the velocities and the forces and it is seen that all quantities diverge exponentially from machine accuracy up to a certain behavior at long times, which is shown in the inset. To understand the long time behavior, $\gamma_x(t)$ can be written as average property

$$\begin{aligned}\gamma_x(t) &= \langle (x(t) - x(0) - \tilde{x}(t) + x(0))^2 \rangle & (87)\\ &= \langle |x(t) - x(0)|^2 \rangle + \langle |\tilde{x}(t) - x(0)|^2 \rangle & (88)\\ &\quad -2\langle x(t)\tilde{x}(t) \rangle + 2\langle x(0)\tilde{x}(t) \rangle + 2\langle x(t)x(0) \rangle - 2\langle x(0)^2 \rangle\end{aligned}$$

In the second equation the first two terms are mean square displacements of x in the two systems (note that $\tilde{x}(0) = x(0)$ since the same starting configurations are used), the next term is a cross correlation between the systems. This will vanish if the systems become independent of each other. The next two systems consist of auto-correlation functions of x in each system. For long times they will also decrease to zero. Finally, the last term gives a constant offset which does not depend on time. Therefore the long time behavior will be governed for coordinates, momenta and forces by

$$\lim_{t\to\infty} \gamma_q(t) = 2\langle |\mathbf{q}(t) - \mathbf{q}(0)|^2 \rangle = 12Dt \quad (89)$$

$$\lim_{t\to\infty} \gamma_p(t) = 2\langle \mathbf{p}(t)^2 \rangle = mk_BT \quad (90)$$

$$\lim_{t\to\infty} \gamma_f(t) = 2\langle \mathbf{F}(t)^2 \rangle = 2(\nabla\mathcal{W})^2 \quad (91)$$

where D is the diffusion coefficient, T the temperature and $\mathcal{W}$ the potential of mean force.

That the divergent behavior of neighbored trajectories is a system dependent property is shown in Figure 10 where results for Lennard-Jones systems at different temperatures are shown.

In conclusion, the individual trajectories of a physical complex system will end up at different places in phase space when introducing round off errors or small perturbations. Round off errors cannot be avoided with simple floating-point arithmetic (only discrete calculations are able to avoid round off errors; but then the physical problem is transformed into a different space). Since one cannot say anything about a *true* summation order, the location in phase space cannot have an absolute meaning. Therefore, the solution to come out of this dilemma is to interpret the phase space location as a *possible* and *allowed* realization of the system, which makes it necessary, however, to average over a lot of possible realizations.

3.5 Accuracy

For an integrator of order $p \geq 1$, the local error may be written as an upper bound[8]

$$\|\Phi_{\delta t,\mathcal{H}} - \phi_{\delta t}\| \leq M\delta t^{p+1} \quad (92)$$

where $M > 0$ is a constant, $\Phi_{\delta t,\mathcal{H}}$ is the exact and $\phi_{\delta t}$ the numerical flow of the system. The global error, i.e. the accumulated error for larger times, is thereby bound for stable

methods by[8]

$$\|\Gamma(t_n) - \Gamma_n\| \leq K\left(e^{Lt_n} - 1\right)\, \delta t^p \qquad , \qquad t_n = n\delta t \tag{93}$$

where $K > 0$ is a constant, $L > 0$ the Lipschitz constant, $\Gamma(t_n) = (\mathbf{p}(t_n), \mathbf{q}(t_n))$ the exact and $\Gamma_n = (\mathbf{p}_n, \mathbf{q}_n)$ the numerically computed trajectory at time t_n. This estimate gives of course not too much information for $Lt_n 1$ unless δt is chosen very small. Nevertheless, qualitatively this estimate shows a similar exponential divergent behavior of numerical and exact solution for a numerical scheme, as was observed in Section 3.4.

A different approach to the error behavior of a numerical scheme is backward error analysis, first mentioned in Ref.[128] in the context of differential equations. The idea is to consider the numerical solution of a given scheme as the exact solution of a modified equation. The comparison of the original and the modified equation then gives qualitative insight into the long time behavior of a given scheme.

It is assumed that the numerical scheme can be expressed as a series of the form

$$\phi_{\delta t}(\Gamma_n) = \Gamma_n + \delta t f(\Gamma) + \delta t^2 g_2(\Gamma) + \delta t^3 g_3(\Gamma) \pm \dots \tag{94}$$

where the g_i are known coefficients and for consistency of the differential equation it must hold

$$f(\Gamma) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \nabla_p \\ \nabla_q \end{pmatrix} \mathcal{H}(\mathbf{p}, \mathbf{q}) \tag{95}$$

On the other hand it is assumed that there exists a modified differential equation of the form

$$\frac{d}{dt}\tilde{\Gamma} = f(\tilde{\Gamma}) + \delta t f_2(\tilde{\Gamma}) + \delta t^2 f_3(\tilde{\Gamma}) + \dots \tag{96}$$

where $\tilde{\Gamma}$ will be equivalent to the numerically obtained solution. In order to construct the modified equation, the solution of Eq. (96) is Taylor expanded, i.e.

$$\begin{aligned}
\tilde{\Gamma}(t+\delta t) = \tilde{\Gamma}(t) &+ \delta t \left(f(\tilde{\Gamma}) + \delta t f_2(\tilde{\Gamma}) + \delta t^2 f_3(\tilde{\Gamma}) + \dots\right) \\
&+ \frac{\delta t^2}{2!}\left(f'(\tilde{\Gamma}) + \delta t f_2'(\tilde{\Gamma}) + \dots\right)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\left(f(\tilde{\Gamma}) + \delta t f_2(\tilde{\Gamma}) + \dots\right) \\
&+ \frac{\delta t^3}{3!}\Bigg\{\left(f''(\tilde{\Gamma}) + \delta t f_2''(\tilde{\Gamma}) + \dots\right)\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\left(f(\tilde{\Gamma}) + \delta t f_2(\tilde{\Gamma}) + \dots\right)\right)^2 \\
&\quad + \left(f'(\tilde{\Gamma}) + \delta t f_2'(\tilde{\Gamma}) + \dots\right)\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\left(f'(\tilde{\Gamma}) + \delta t f_2'(\tilde{\Gamma}) + \dots\right)\right) \\
&\quad \times \left(f(\tilde{\Gamma}) + \delta t f_2(\tilde{\Gamma}) + \dots\right)\Bigg\} \\
&+ \dots
\end{aligned} \tag{97}$$

The procedure to construct the unknown functions f_i proceeds in analogy to perturbation theory, i.e. coefficients with same powers of δt are collected which leads to a recursive scheme to solve for all unknowns.

To give an example the Lennard-Jones oscillator is considered, i.e. a particle performing stable motions in negative part of a Lennard-Jones potential. As was observed already for the harmonic oscillator, the Explicit Euler method will gain energy during the time,

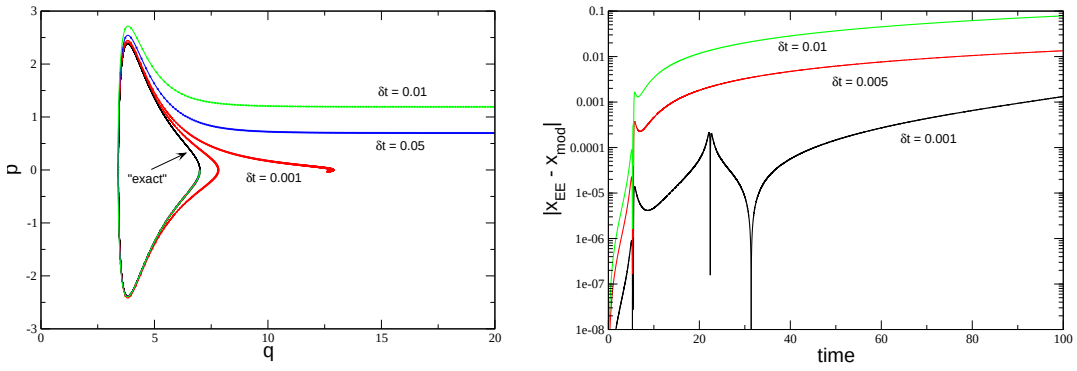


Figure 11. Phase space trajectories of the Lennard-Jones oscillator calculated with the Explicit Euler scheme and different time steps of integration. The *exact* solution (numerical solution of a high order composition scheme with small time step) is shown as a reference - it forms closed orbits. Superimposed to the solutions are results, obtained with a Velocity Verlet scheme, applied to the modified equations, Eqs. (98,99). The right figure shows the differences in coordinates between the calculation with Explicit Euler scheme applied to Lennard-Jones oscillator and Velocity Verlet applied to the modified equation, $|\mathbf{q}_{EE}(t) - \mathbf{q}_{mod}(t)|$.

i.e. the particle will increase kinetic energy which finally will lead to an escape of the Lennard-Jones potential well. Solving for the modified equation of the Explicit Euler, one gets as a first correction

$$\dot{\mathbf{q}} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}} + \frac{\delta t}{2} \frac{\partial \mathcal{H}}{\partial \mathbf{q}} \tag{98}$$

$$\dot{\mathbf{p}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{q}} + \frac{\delta t}{2} \mathbf{p} \frac{\partial^2 \mathcal{H}}{\partial \mathbf{p}^2} \tag{99}$$

Figure 11 shows results for the integration of equations of motion with the Explicit Euler scheme. Different time steps for integration were applied which show a faster escape from a stable orbit with increasing time step. Also plotted in the same figure is the solution of the modified equations with a high order symplectic scheme, which can be considered as *exact* on these time scales. It is found that the trajectories more or less coincide and cannot be distinguished by eye. A more quantitative analysis (Figure 11) shows that for relatively long times the solution is rather well approximated by the modified equation, although with increasing time the differences between solutions become more pronounced. This means that for longer times it would be necessary to include more terms of higher order in δt into the modified equation. It should be mentioned that, in general, the series expansion of the modified equation diverges.

4 Simulating in Different Ensembles

In MD simulations it is possible to realize different types of thermodynamic ensembles which are characterized by the control of certain thermodynamic quantities. If one knows how to calculate a thermodynamic quantity, e.g. the temperature or pressure, it is often possible to formulate an algorithm which fixes this property to a desired value. However, it is

not always clear whether this algorithm describes the properties of a given thermodynamic ensemble.

One can distinguish four different types of control mechanisms:

Differential control : the thermodynamic quantity is fixed to the prescribed value and no fluctuations around an average value occur.

Proportional control : the variables, coupled to the thermodynamic property f, are corrected in each integration step through a coupling constant towards the prescribed value of f. The coupling constant determines the strength of the fluctuations around $\langle f \rangle$.

Integral control : the system's Hamiltonian is extended and variables are introduced which represent the effect of an external system which fix the state to the desired ensemble. The time evolution of these variables is determined by the equations of motion derived from the extended Hamiltonian.

Stochastic control : the values of the variables coupled to the thermodynamic property f are propagated according to modified equations of motion, where certain degrees of freedom are additionally modified stochastically in order to give the desired mean value of f.

In the following, different statistical ensembles are presented and all methods will be discussed via examples.

4.1 The Microcanonical Ensemble

The microcanonical ensemble (NVE) may be considered as the *natural* ensemble for molecular dynamics simulations (as it is the canonical ensemble (NVT) for Monte Carlo simulations). If no time dependent external forces are considered, the system's Hamiltonian is constant, implying that the system's dynamics evolves on a constant energy surface. The corresponding probability density in phase space is therefore given by

$$\rho(\mathbf{q}, \mathbf{p}) = \delta(\mathcal{H}(\mathbf{q}, \mathbf{p}) - E) \tag{100}$$

In a computer simulation this theoretical condition is generally violated, due to limited accuracy in integrating the equations of motion and due to roundoff errors resulting from a limited precision of number representation. In Ref.[129] a numerical experiment was performed showing that tiny perturbations of the initial positions of a trajectory are doubled about every picosecond. This would mean even for double precision arithmetic that after about $50\ ps$ roundoff errors will be dominant[117]. This is, however, often not a too serious restriction, since most time correlation functions drop to zero on a much shorter time scale. Only for the case where long time correlations are expected one does have to be very careful in generating trajectories.

4.2 The Canonical Ensemble

The simplest extension to the microcanonical ensemble is the canonical one (N,V,T), where the number of particles, the volume and the temperature are fixed to prescribed values. The temperature T is, in contrast to N and V, an intensive parameter. The extensive counterpart would be the kinetic energy of the system. In the following, different control mechanisms, introduced in Sec. 4 are described.

4.2.1 The Differential Thermostat

Different methods were proposed to fix the temperature to a fixed value during a simulation without allowing fluctuations of T. The first method was introduced by Woodcock[130], where the velocities were scaled according to $\mathbf{p}_i \rightarrow \sqrt{T_0/T}\mathbf{p}_i$, where T_0 is the reference temperature and T the actual temperature, calculated from the velocity of the particles. This method leads to discontinuities in the momentum part of the phase space trajectory due to the rescaling procedure.

An extension of this method implies a constraint of the equations of motion to keep the temperature fixed[131–133]. The principle of least constraint by Gauss states that a force added to restrict a particle motion on a constraint hypersurface should be normal to the surface in a realistic dynamics. From this principle the equations of motion are derived

$$\frac{\partial \mathbf{q}_i}{\partial t} = \mathbf{p}_i \tag{101}$$

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial V}{\partial \mathbf{q}_i} - \zeta \mathbf{p}_i \tag{102}$$

where ζ is a constraint force term, calculated as

$$\zeta = -\frac{\sum_{i=1}^{N} \frac{\mathbf{p}_i}{m_i} \frac{\partial V}{\partial \mathbf{q}_i}}{\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{m_i}} \tag{103}$$

Since the principle of least constraint by Gauss is used, this algorithm is also called *Gaussian thermostat*. It may be shown for this method that the configurational part of the phase space density is of canonical form, i.e.

$$\rho(\mathbf{q}, \mathbf{p}) = \delta(T - T_0)\, e^{-\beta U(\mathbf{q})} \tag{104}$$

4.2.2 The Proportional Thermostat

The proportional thermostat tries to correct deviations of the actual temperature T form the prescribed one T_0 by multiplying the velocities by a certain factor λ in order to move the system dynamics towards one corresponding to T_0. The difference with respect to the differential control is that the method allows for fluctuations of the temperature, thereby not fixing it to a constant value. In each integration step it is insured that the T is corrected to a value more close to T_0. A thermostat of this type was proposed by Berendsen et al.[134,135] who introduced *weak coupling methods to an external bath*. The weak coupling thermostat was motivated by the minimization of local disturbances of a stochastic thermostat while keeping the global effects unchanged. This leads to a modification of the momenta $\mathbf{p}_i \rightarrow \lambda \mathbf{p}_i$, where

$$\lambda = \left[1 + \frac{\delta t}{\tau_T}\left(\frac{T_0}{T} - 1\right)\right]^{\frac{1}{2}} \tag{105}$$

The constant τ_T, appearing in Eq.105, is a so called coupling time constant which determines the time scale on which the desired temperature is reached. It is easy to show that the

proportional thermostat conserves a Maxwell distribution. However, the method cannot be mapped onto a specific thermodynamic ensemble. In Ref.[136] the phase space distribution could be shown to be

$$\rho(\mathbf{q},\mathbf{p}) = f(\mathbf{p})\, e^{-\beta\left(U(\mathbf{q})-\alpha\beta\delta U(\mathbf{q})^2/3N\right)} \tag{106}$$

where $\alpha \simeq (1 - \delta E/\delta U)$ and $\delta U,\ \delta E$ are the mean fluctuations of the potential and total energy. $f(\mathbf{p})$ is in general an unknown function of the momenta, so that the full density cannot be determined. For $\alpha = 0$, which corresponds in Eq.105 to $\tau_T = \delta t$, the fluctuations in the kinetic energy vanish and Eq.106 reduces to Eq.104, i.e. it represents the canonical distribution. The other extreme of $\tau_T \rightarrow \infty$ corresponds to an isolated system and the energy should be conserved, i.e. $\delta E = \delta K + \delta U = 0$ and $\alpha = 1$. In this case, Eq.106 corresponds to the microcanonical distribution[136]. Eq.106 may therefore be understood as an interpolation between the canonical and the microcanonical ensemble.

4.2.3 The Stochastic Thermostat

In the case of a stochastic thermostat, all or a subset of the degrees of freedom of the system are subject to collisions with *virtual* particles. This method can be motivated by a Langevin stochastic differential equation which describes the motion of a particle due to the thermal agitation of a heat bath

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial U}{\partial \mathbf{q}_i} - \gamma \mathbf{p}_i + \mathbf{F}^+ \tag{107}$$

where γ is a friction constant and $\mathbf{F}^+$ a Gaussian random force. The amplitude of $\mathbf{F}^+$ is determined by the second fluctuation dissipation theorem

$$\langle \mathbf{F}_i^+(t_1)\mathbf{F}_j^+(t_2)\rangle = 2\gamma k_B T \delta_{ij}\delta(t_1 - t_2) \tag{108}$$

A larger value for γ will increase thermal fluctuations, while $\gamma = 0$ reduces to the microcanonicle ensemble. This method was applied to molecular dynamics in Ref.[137]. A more direct way was followed in Refs.[138,139] where particles collide occasionally with virtual particles from a Maxwell distribution corresponding to a temperature T_0 and after collisions loose their memory completely, i.e. the motion is totally randomized and the momenta become discontinuous. In order not to disturb the phase space trajectory too much, the collision frequency has to be chosen not too high. Since a large collision frequency will lead to a strong loss of the particle's memory, it will lead to a fast decay of dynamic correlation functions[140]. The characteristic decay time of correlation functions should therefore be a measure for the collision time. It was proved for the stochastic thermostat[138] that it leads to a canonical distribution function.

A slightly different method which is able to control the coupling to an external bath was suggested in Refs.[141,142]. In this approach the memory of the particle is not completely destroyed but the new momenta are chosen to be

$$\mathbf{p}_{i,n} = \sqrt{1-\alpha^2}\, \mathbf{p}_{i,o} + \alpha\, \mathbf{p}_r \tag{109}$$

where $\mathbf{p}_r$ is chosen a momentum, drawn from a Maxwell distribution corresponding to T_0. Similar to the proportional thermostat, the parameter α may be tuned to give distributions ranging from the microcanonical to the canonical ensemble.

4.2.4 The Integral Thermostat

The integral method is also often called *extended system method* as it introduces additional degrees of freedom into the system's Hamiltonian for which equations of motion can be derived. They are integrated in line with the equations for the spatial coordinates and momenta. The idea of the method invented by Nosé[143,144], is to reduce the effect of an external system acting as heat reservoir to keep the temperature of the system constant, to one additional degree of freedom. The thermal interactions between a heat reservoir and the system result in a change of the kinetic energy, i.e. the velocity of the particles in the system. Formally it may therefore be expressed a scaling of the velocities. Nosé introduced two sets of variables: real and so called virtual ones. The virtual variables are consistently derived from a Sundman transformation[145] $d\tau/dt = s$, where τ is a virtual time and s is a resulting scaling factor, which is treated as dynamical variable. The transformation from virtual to real variables is then performed as

$$\mathbf{p}_i = \boldsymbol{\pi}_i s \qquad , \qquad \mathbf{q}_i = \boldsymbol{\rho}_i \tag{110}$$

The introduction of the *effective mass*, M_s, connects also a momentum to the additional degree of freedom, π_s. The resulting Hamiltonian, expressed in virtual coordinates reads

$$\mathcal{H}^* = \sum_{i=1}^{N} \frac{\boldsymbol{\pi}_i^2}{2m_i s^2} + U(\boldsymbol{\rho}) + \frac{\pi_s^2}{2M_s} + g k_B T \ln s \tag{111}$$

where $g = 3N + 1$ is the number of degrees of freedom (system of N free particles). The Hamiltonian in Eq.111 was shown to lead to a probability density in phase space, corresponding to the canonical ensemble.

The equations of motion drawn from this Hamiltonian are

$$\frac{\partial \boldsymbol{\rho}_i}{\partial \tau} = \frac{\boldsymbol{\pi}_i}{s^2} \tag{112}$$

$$\frac{\partial \boldsymbol{\pi}_i}{\partial \tau} = -\frac{\partial U(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}_i} \tag{113}$$

$$\frac{\partial s}{\partial \tau} = \frac{\pi_s}{M_s} \tag{114}$$

$$\frac{\partial \pi_s}{\partial \tau} = \frac{1}{s^3} \sum_{i=1}^{N} \frac{\boldsymbol{\pi}_i^2}{m_i} - \frac{g k_B T}{s} \tag{115}$$

If one transforms these equations back into real variables, it is found[146] that they can be simplified by introducing the new variable $\zeta = \partial s/\partial t = s p_s/M_s$ (p_s is *real* momentum connected to the heat bath)

$$\frac{\partial \mathbf{q}_i}{\partial t} = \frac{\mathbf{p}_i}{m_i} \tag{116}$$

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial U(\mathbf{q})}{\partial \mathbf{q}_i} - \zeta \mathbf{p}_i \tag{117}$$

$$\frac{\partial \ln s}{\partial t} = \zeta \tag{118}$$

$$\frac{\partial \zeta}{\partial t} = \frac{1}{M_s} \left(\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{m_i} - g k_B T \right) \tag{119}$$

These equations describe the so called Nosé-Hoover thermostat.

4.3 The Constant-Pressure Constant-Enthalpy Ensemble

In order to control the pressure in an MD simulation cell, it is necessary to allow for volume variations. A simple picture for a constant pressure system is a box the walls of which are coupled to a piston which controls the pressure. In contrast to the case where the temperature is controlled, no coupling to the dynamics of the particles (timescales) is performed but the length scales of the system will be modified. In the following, different algorithms are described for a constant pressure ensemble. The conserved quantity will not be the system's energy, since there will be an energy transfer to or from the *external* system (piston etc.), but the enthalpy H will be constant. In line with the constant temperature methods there are also differential, proportional, integral and stochastic methods to achieve a constant pressure situation in simulations. The differential method, however, is not discussed here, since there are problems with that method related to the *correct initial* pressure[147, 148]. A scheme for the calculation of the pressure in MD simulations for various model systems is given in the appendix.

4.3.1 The Proportional Barostat

The proportional thermostat in Sec. 4.2.2 was introduced as an extension for the equation of the momentum, since it couples to the kinetics of the particles. Since the barostat acts on a volume change, which may be expressed in a scaling of particles' positions, a phenomenological extension for the equation of motion of the coordinates may be formulated[134]

$$\frac{\partial \mathbf{q}_i}{\partial t} = \frac{\mathbf{p}_i}{m_i} + \alpha \mathbf{q}_i \tag{120}$$

while a change in volume is postulated as

$$\dot{V} = 3\alpha V \tag{121}$$

A change in pressure is related to the isothermal compressibility κ_T

$$\dot{P} = -\frac{1}{\kappa_T V}\frac{\partial V}{\partial t} = -\frac{3\alpha}{\kappa_T} \tag{122}$$

which is approximated as

$$\frac{(P_0 - P)}{\tau_P} = -\frac{3\alpha}{\kappa_T} \tag{123}$$

and therefore Eq.120 can be written as

$$\frac{\partial \mathbf{q}_i}{\partial t} = \frac{\mathbf{p}_i}{m_i} - \frac{\kappa_T}{3\tau_P}(P_0 - P) \tag{124}$$

which corresponds to a scaling of the boxlength and coordinates $\mathbf{q} \rightarrow s\mathbf{q}$ and $L \rightarrow sL$, where

$$s = 1 - \frac{\kappa_T \delta t}{3\tau_P}(P_0 - P) \tag{125}$$

The time constant τ_P was introduced into Eq.123 as a characteristic timescale on which the system pressure will approach the desired pressure P_0. It also controls the strength of the coupling to the barostat and therefore the strength of the volume/pressure fluctuations. If the isothermal compressibility, κ_T, is not known for the system, the constant $\tau_P' = \tau_P/\kappa_T$ may be considered as a phenomenological coupling time which can be adjusted to the system under consideration. As for the proportional thermostat, a drawback for this method is that the statistical ensemble is not known. In analog to the thermostat, it may be assumed to *interpolate* between the microcanonical and the constant-pressure/constant-enthalpy ensemble, depending on the coupling constant τ_P.

4.3.2 The Integral Barostat

In line with the integral thermostat one can introduce a new degree freedom into the systems Hamiltonian which controls volume fluctuations. This method was first proposed by Andersen[138]. The idea is to include the volume as an additional degree of freedom and to write the Hamiltonian in a scaled form, where lengths are expressed in units of the boxlength $L = V^{1/3}$, i.e. $\mathbf{q}_i = L\ \boldsymbol{\rho}_i$ and $\mathbf{p}_i = L\ \boldsymbol{\pi}_i$. Since L is also a dynamical quantity, the momentum is not related to the simple time derivative of the coordinates but $\partial_t \mathbf{q}_i = L\ \partial_t \boldsymbol{\rho}_i + \boldsymbol{\rho}_i\ \partial_t L$. The extended system Hamiltonian is then written as

$$\mathcal{H}^* = \frac{1}{V^{2/3}} \sum_{i=1}^{N} \frac{\boldsymbol{\pi}_i}{2m_i} + U(V^{1/3}\boldsymbol{\rho}) + P_{ex}V + \frac{\boldsymbol{\pi}_V^2}{2M_V} \tag{126}$$

where P_{ex} is the prescribed external pressure and $\boldsymbol{\pi}_V$ and M_V are a momentum and a mass associated with the fluctuations of the volume.

The equations of motion which are derived from this Hamiltonian are

$$\frac{\partial \boldsymbol{\rho}_i}{\partial t} = \frac{1}{V^{2/3}} \frac{\boldsymbol{\pi}_i}{m_i} \tag{127}$$

$$\frac{\partial \boldsymbol{\pi}_i}{\partial t} = \frac{\partial U(V^{1/3}\boldsymbol{\rho})}{\partial \boldsymbol{\rho}_i} \tag{128}$$

$$\frac{\partial V}{\partial t} = \frac{\boldsymbol{\pi}_V}{M_V} \tag{129}$$

$$\frac{\partial \boldsymbol{\pi}_V}{\partial t} = \frac{1}{3V} \left(\frac{1}{V^{2/3}} \sum_{i=1}^{N} \frac{\boldsymbol{\pi}_i}{m_i} - V^{1/3} \boldsymbol{\rho}_i \frac{\partial U(\mathbf{q})}{\partial \mathbf{q}_i} \right) \tag{130}$$

A transformation to real variables then gives

$$\frac{\partial \mathbf{q}_i}{\partial t} = \frac{\mathbf{p}_i}{m_i} + \frac{1}{3V} \frac{\partial V}{\partial t} \mathbf{q}_i \tag{131}$$

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial U(\mathbf{q})}{\partial \mathbf{q}_i} - \frac{1}{3V} \frac{\partial V}{\partial t} \mathbf{p}_i \tag{132}$$

$$\frac{\partial V}{\partial t} = \frac{\boldsymbol{p}_V}{M_V} \tag{133}$$

$$\frac{\partial \boldsymbol{p}_V}{\partial t} = \underbrace{\frac{1}{3V} \left(\sum_{i=1}^{N} \frac{\mathbf{p}_i}{m_i} - \mathbf{q}_i \frac{\partial U(\mathbf{q})}{\partial \mathbf{q}_i} \right)}_{P} - P_{ex} \tag{134}$$

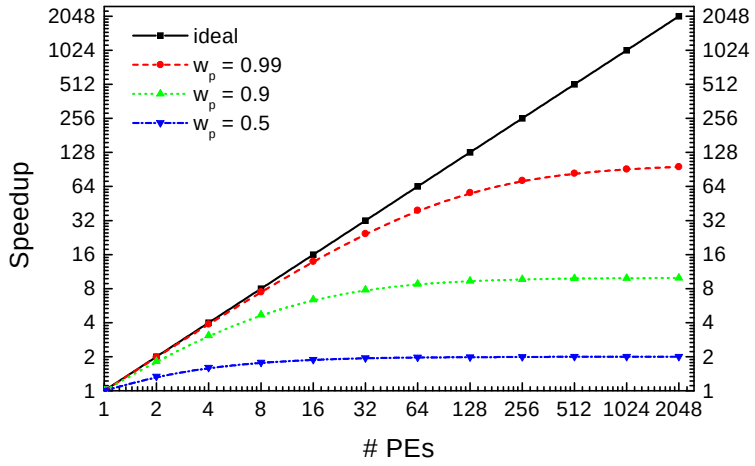


Figure 12. The ideal speedup for parallel applications with 50%, 90%, 99% and 100% (ideal) parallel work as a function of the number of processors.

In the last equation the term in brackets corresponds to the pressure, calculated from the virial theorem. The associated volume force, introducing fluctuations of the box volume is therefore controlled by the internal pressure, originating from the particle dynamics and the external pressure, P_{ex}.

5 Parallel Molecular Dynamics

With the advent of massively parallel computers, where thousands of processors may work on a single task, it has become possible to increase the size of the numerical problems considerably. As has been already mentioned in Sec.1 it is in principle possible to treat multi-billion particle systems. However, the whole success of parallel computing strongly depends both on the underlying problem to be solved and the optimization of the computer program. The former point is related to a principle problem which is manifested in the so called Amdahl's law[149]. If a problem has inherently certain parts which can be solved only in serial, this will give an upper limit for the parallelization which is possible. The speedup σ, which is a measure for the gain of using multiple processors with respect to a single one, is therefore bound

$$\sigma = \frac{N_p}{w_p + N_p w_s}. \tag{135}$$

Here, N_p is the number of processors, w_p and w_s is the amount of work, which can be executed in parallel and in serial, i.e. $w_p + w_s = 1$. From Eq.135 it is obvious that the maximum efficiency is obtained when the problem is completely parallelizable, i.e. $w_p = 1$ which gives an N_p times faster execution of the program. In the other extreme, when $w_s = 1$ there is no gain in program execution at all and $\sigma = 1$, independent of N_p. In Fig.12 this limitation is illustrated for several cases, where the relative amount for the serial work was modified. If the parallel work is 50%, the maximum speedup is bound to $\sigma = 2$.

If one aims to execute a program on a real massively parallel computer with hundreds or thousands of processors, the problem at hand must be inherently parallel for 99.99...%. Therefore, not only big parallel computers guarantee a fast execution of programs, but the problem itself has to be chosen properly.

Concerning MD programs there are only a few parts which have to be analysed for parallelization. As was shown, an MD program consists essentially of the force routine, which costs usually more than 90% of the execution time. If one uses neighbor lists, these may be also rather expensive while reducing the time for the force evaluation. Other important tasks are the integration of motion, the parameter setup at the beginning of the simulation and the file input/output (I/O). In the next chapter it will be shown how to parallelize the force routine. The integrator may be naturally parallelized, since the loop over N particles may be subdivided and performed on different processors. The parameter setup has either to be done in serial so that every processor has information about relevant system parameters, or it may be done in parallel and information is distributed from every processor via a broadcast. The file I/O is a more complicated problem. The message passing interface MPI I does not offer a parallel I/O operation. In this case, if every node writes some information to the same file there is, depending on the configuration of the system, often only one node for I/O, to which internally the data are sent from the other nodes. The same applies for reading data. Since on this node the data from/for the nodes are written/read sequentially, this is a serial process which limits the speedup of the execution. The new MPI II standard offers parallel read/write operations, which lead to a considerable efficiency gain with respect to MPI. However, the efficiency obtained depend strongly on the implementation on different architectures. Besides MPI methods, there are other libraries, which offer more efficient parallel I/O with respect to native programming. To name a few, there are PnetCDF[150,151], an extension towards parallelism of the old *network Common Data Form*, netCDF-4[152,153], which is in direct line of netCDF development, which now has parallel functionality and which is built on top of MPI-I/O, or SIONlib, a recently developed high performance library for parallel I/O[154].

Another serious point is the implementation into the computer code. A problem which is inherently 100% parallel will not be solved with maximum speed if the program is not 100% mapped onto this problem. Implementation details for parallel algorithms will be discussed in the following sections. Independent of the implementation of the code, Eq.135 gives only an upper theoretical limit which will only be reached in very rare cases. For most problems it is necessary to communicate data from one processor to another or even to all other processors in order to take into account data dependencies. This implies an overhead which depends on the latency and the bandwidth of the interprocessor network, which strongly depends on the hardware.

5.1 Domain Decomposition

The principle of spatial decomposition methods is to assign geometrical domains to different processors. This implies that particles are no longer bound to a certain processor but will be transfered from one PE to another, according to their spatial position. This algorithm is especially designed for systems with short range interactions or to any other algorithm where a certain cut-off in space may be applied. Since neighbored processors contain all relevant data needed to compute forces on particles located on a given PE,

this algorithm avoids the problem of global communications. Given that the range of interaction between particles is a cut-off radius of size R_c, the size, D of the domains is preferentially chosen to be $D > R_c$, so that only the $3^d - 1$ neighbored processors have to communicate data (d is the dimension of the problem). Whether this can be fulfilled depends on the interplay between size of the system and the numbers of processors. If a small system is treated with a large number of processors, the domains will be small and $D < R_c$. In this case not only the next but also the second or even higher order neighbor PEs have to send their coordinates to a given PE. For simplicity we assume here $D > R_c$. Algorithms, which treat efficiently the general case were developed recently[155–157].

The algorithm then works as follows. Particles are distributed in the beginning of the simulation to a geometrical region. The domains are constructed to have a rather homogeneous distribution of particles on each processor, e.g. for homogeneous bulk liquids the domains can be chosen as equally sized cuboids which fill the simulation box. In order to calculate forces between particles on different processors, coordinates of the so called *boundary particles* (those which are located in the outer region of size $R_b \geq R_c$ of the domains) have to be exchanged. Two types of lists are constructed for this purpose. The one contains all particle indices, which have left the local domain and which have consequently to be transferred to the neighbored PE. The other one contains all particle indices, which lie in the outer region of size R_b of a domain. The first list is used to update the particles' *address*, i.e. all information like positions, velocities, forces etc. are sent to the neighbored PE and are erased in the old domain. The second list is used to send temporarily position coordinates which are only needed for the force computation. The calculation of forces then operates in two steps. First, the forces due to local particles are computed using Newton's 3rd law. In a next step, forces due to the boundary particles are calculated. The latter forces are thus calculated twice: on the local PE and the neighbored PE. This extra computation has the advantage that there is no communication step for forces. A more elaborate scheme has nevertheless been proposed which includes also Newton's 3rd law for the boundary particles and thus the communication of forces[158, 159]. Having finished the evaluation of forces, the new positions and velocities are evaluated only for local particles.

A naive method would require $3^d - 1$ send/receive operations. However, this may be reduced to $2\log_d(3^d - 1)$ operations with a similar tree-like method. The method is described here for the case of $d = 2$. It may be generalized rather easily. The 4 processors, located directly at the edges of a given one are labeled as left/right and up/down. Then in a first step, information is sent/received to/from the left and the right PE, i.e. each processor now stores the coordinates of three PEs (including local information). The next step proceeds in sending/receiving the data to the up and down PEs. This step finishes already the whole communication process.

The updating process is not necessarily done in each time step. If the width of the boundary region is chosen as $R_b = R_c + \delta r$, it is possible to trigger the update automatically via the criterion $\max(|\mathbf{x}(t_0 + t) - \mathbf{x}(t_0)|) \leq \delta r$, which is the maximum change in distance of any particle in the system, measured from the last update.

A special feature of this algorithm is the fact that it shows a theoretical superlinear speed-up if Verlet neighbor lists are used. The construction of the Verlet list requires $N'(N' - 1)/2 + N'\delta N$ operations, where δN is the number of boundary particles and N' is the number of particles on a PE. If the number of PEs is increased as twice as large,

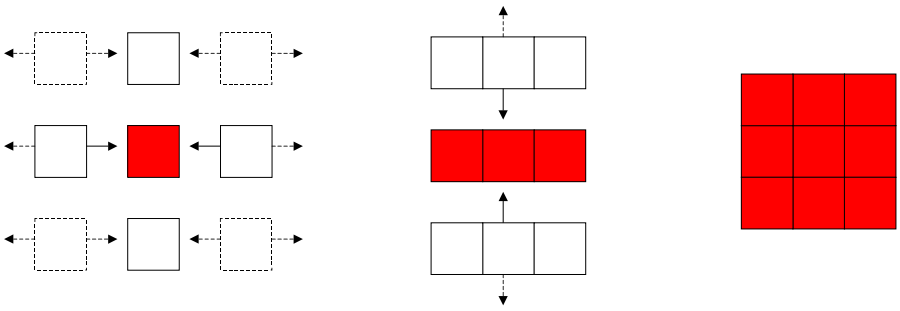


Figure 13. Communication pattern for the domain decomposition algorithm in 2 dimensions.

there are $N'/2$ particles on each processor which therefore requires $N'/2(N'/2-1)/2 + N'/2\delta N$ operations. If $N' \gg \delta N$ and ${N'}^2 \gg N'$ one gets a speed-up factor of ≈ 4!

5.2 Performance Estimations

In order to estimate the performance of the different algorithms on a theoretical basis it is useful to extend the ideal Amdahl's law to a more realistic case. The ideal law only takes into account the degree of parallel work. From that point of view all parallel algorithms for a given problem should work in the same way. However the communication between the processors is also a limiting factor in parallel applications and so it is natural to extend Amdahl's law in the following way

$$\sigma = \frac{1}{w_p/N_p + w_s + c(N_p)} \tag{136}$$

where $c(N_p)$ is a function of the number of processors which will characterize the different parallel algorithms. The function will contain both communication work, which depends on the bandwidth of the network and the effect of the latency time, i.e. how fast the network responds to the communication instruction. The function $c(N_p)$ expresses the relative portion of communication with respect to computation. Therefore it will depend in general also on the number of particles which are simulated.

In the following a model analysis for the domain decomposition algorithm is presented. It is assumed that the work is strictly parallel, i.e. $w_p = 1$.

Spatial decomposition algorithm is based on local communication. As was described in Sec.5.1, only six communication steps are required to distribute the data to neighbored PEs. Therefore the latency time part is constant whereas the amount of data to be sent and consequently the communication part is decreased with larger N_p. The communication function reads therefore

$$c(N_p) = f(N_p)\left(\lambda + \frac{\chi}{N_p^{2/3}}\right) \quad , \quad f(N_p) = \begin{cases} 0 & N_p = 1 \\ 2 & N_p = 2 \\ 4 & N_p = 4 \\ 6 & N_p \leq 8 \end{cases} \tag{137}$$

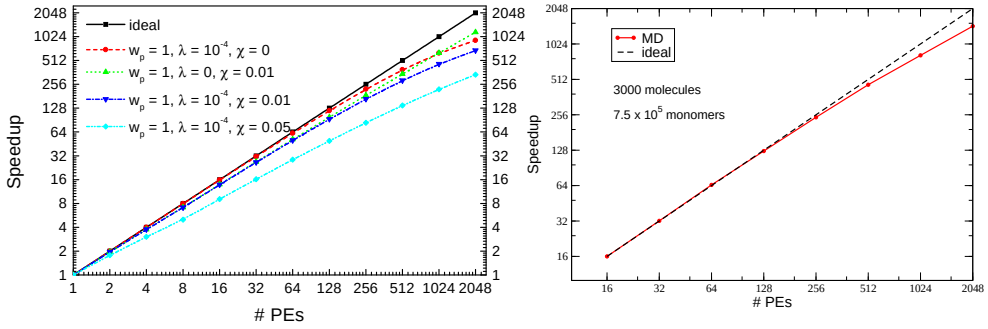


Figure 14. Left: Estimation of realistic speedup curves if one includes the latency time and bandwidth of the processor interconnect. It is assumed that the problem can potentially be parallelized 100%. Different parameter values are compared for the latency time λ and bandwidth χ for a local nearest neighbor communications. The ideal curve neglects communication completely. Right: Realistic benchmark for a domain decomposition program, simulating a system consisting of 3000 polymers with 250 monomers each.

Here the function $f(N_p)$ was introduced to cover also the cases for small numbers of PEs, where a data exchange is not necessary in each spatial direction. As seen from Fig.14 the speedup curves are nearly linear with a slightly smaller slope than unity. However, for very large numbers of PEs the curves will also flatten. Nevertheless, the local communication model provides the best speedup behavior from all parallel algorithms and seems to be best suited for large parallel architectures.

Remark Note that the local communication model in its present form is only valid for short range interaction potentials. If the potential is longer ranged than one spatial domain, the function $f(N_p)$ has to be modified. For long range interactions, all-to-all communications are generally required. In that case the tree-method may be mostly preferred.

This theoretical analysis demonstrates the importance of a fast interconnect between processors for the case of molecular dynamics simulations. Not included in the communication function $c(N_p)$ is the bandwidth function of the network. This, however, will only slightly change Fig.14.

5.3 Comparison with Simulation

In order to verify the theoretical model, one may perform real MD simulations for model systems, which are as large as the principal features, appearing in the analysis are fulfilled. This includes that domains are large enough in order to restrict particle interactions to neighbored domains and to have a nearly homogenous particle distribution, which avoids unbalanced computational work on the processors.

In the current case, the program MP2C[160] was used, which implements both a mesoscopic solvent method, based on the Muli-Particle-Collision (MPC) dynamics and a molecular dynamics part. The program is based on a domain decomposition approach and allows to couple MD and MPC simulations or to decouple them, in order to run either MD or MPC in a simulation for e.g. all-atom force-field simulations without hydrodynamic coupling or e.g. fluid dynamics without solvated particles, respectively. In the present case a simulation of a polymer system, consisting of 3000 polymeric chains with 250 monomers each

was simulated. The monomers were coupled within the chain by a harmonic bond potential and the non-bonded part of the potential was set to the repulsive part of a Lennard-Jones potential which was applied to all particle pairs which were not coupled within bonds.

The program was run on an IBM BlueGene/P at Jülich Supercomputing Centre. Fig. 14 shows the scaling up to $N_p = 2048$ processors, which is qualitatively comparable and shows the same behavior as prescribed by the simple model. A better scaling is to be expected, when more particles are simulated, which moves the ratio of communication/computation to smaller values, which reduces the relative overhead in the parallel execution.

References

1. K. Binder and D. Landau. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, Cambridge, 2000.
2. A.R. Leach. *Molecular Modelling - Principles and Applications*. Pearson Education Ltd., Essex, England, 2001.
3. T. Schlick. *Molecular Modeling and Simulation*. Springer, New York, 2002.
4. K. Binder and D.W. Heermann. *Monte Carlo Simulation in Statistical Physics*. Springer, Berlin, 1997.
5. D. Frenkel and B. Smit. *Understanding molecular simulation. From algorithms to applications.* Academic Press, San Diego, 1996.
6. J. M. Haile. *Molecular Dynamics Simulation*. Wiley, New York, 1997.
7. H. Goldstein, Ch. Poole, and J. Safko. *Classical Mechanics*. Addison Wesley, San Francisco, CA, 2002.
8. B. Leimkuhler and S. Reich. *Simulating Hamiltonian Dynamics*. Cambridge University Press, Cambridge, 2004.
9. B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *J. Chem. Phys.*, 27:1208–1209, 1957.
10. B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. I. General method. *J. Chem. Phys.*, 31:459, 1959.
11. J. Roth, F. Gähler, and H.-R. Trebin. A molecular dynamics run with 5.180.116.000 particles. *Int. J. Mod. Phys. C*, 11:317–322, 2000.
12. K. Kadau, T. C. Germann, and P. S. Lomdahl. Large-scale molecular-dynamics simulation of 19 billion particles. *Int. J. Mod. Phys. C*, 15:193, 2004.
13. K. Kadau, T. C. Germann, and P. S. Lomdahl. World record: Large-scale molecular-dynamics simulation of 19 billion particles. Technical Report LA-UR-05-3853, Los Alamos NAtional Laboratory, 2005.
14. K. Kadau, T. C. Germann, and P. S. Lomdahl. Molecular-Dynamics Comes of Age: 320 Billion Atom Simulation on BlueGene/L. *Int. J. Mod. Phys. C*, 17:1755, 2006.
15. T. C. Germann and K. Kadau. Trillion-atom molecular dynamics becomes a reality. *Int. J. Mod. Phys. C*, 19:1315–1319, 2008.
16. P.S. Lomdahl, P. Tamayo, N. Gronbach-Jensen, and D.M. Beazley. In G.S. Ansell, editor, *Proc. Supercomputing 93*, page 520, Los Alamitos CA, 1993. IEEE Computer Society Press.
17. D.M. Beazley and P.S. Lomdahl. *Comput. Phys.*, 11:230, 1997.
18. Y. Duan, L. Wang, and P. A. Kollman. The early stage of folding of villin headpiece subdomain observed in 200-nanosecond fully solvated molecular dynamics simultion.

Proc. Natl. Acad. Sci. USA, 95:9897, 1998.
19. Y. Duan and P. A. Kollman. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282:740, 1998.
20. C. Mura and J.A. McCammon. Molecular dynamics of a κB DNA element: base flipping via cross-strand intercalative stacking in a microsecond-scale simulation. *Nucl. Acids Res.*, 36:4941–4955, 2008.
21. http://www.ccp5.ac.uk/.
22. http://amber.scripps.edu/.
23. http://www.charmm.org.
24. http://www.ks.uiuc.edu/Research/namd/.
25. http://www.emsl.pnl.gov/docs/nwchem/nwchem.html.
26. http://www.gromacs.org.
27. http://www.cs.sandia.gov/ sjplimp/lammps.html.
28. Gregory A. Voth. *Coarse-Graining of Condensed Phase and Biomolecular Systems*. CRC Press, 2008.
29. A. Arkhipov A.Y. Shih, P.L. Freddolino, and K. Schulten. Coarse grained protein-lipid model with application to lipoprotein particles. *J. Phys. Chem. B*, 110:3674–3684, 2006.
30. P.M. Kasson, A. Zomorodian, S. Park, N. Singhal, L.J. Guibas, and V.S. Pande. Persistent voids: a new structural metric for membrane fusion. *Bioinformatics*, 23:1753–1759, 2007.
31. A. J. Stone. Intermolecular potentials. *Science*, 321:787–789, 2008.
32. W. L. Cui, F. B. Li, and N. L. Allinger. *J. Amer. Chem. Soc.*, 115:2943, 1993.
33. N. Nevins, J. H. Lii, and N. L. Allinger. *J. Comp. Chem.*, 17:695, 1996.
34. S. L. Mayo, B. D. Olafson, and W. A. Goddard. *J. Phys. Chem.*, 94:8897, 1990.
35. M. J. Bearpark, M. A. Robb, F. Bernardi, and M. Olivucci. *Chem. Phys. Lett.*, 217:513, 1994.
36. T. Cleveland and C. R. Landis. *J. Amer. Chem. Soc.*, 118:6020, 1996.
37. A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. *J. Amer. Chem. Soc.*, 114:10024, 1992.
38. Z. W. Peng, C. S. Ewig, M.-J. Hwang, M. Waldman, and A. T. Hagler. Derivation of class ii force fields. 4. van der Waals parameters of Alkali metal cations and Halide anions. *J. Phys. Chem.*, 101:7243–7252, 1997.
39. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Amer. Chem. Soc.*, 117:5179–5197, 1995.
40. A. D. Mackerell, J. Wiorkiewiczkuczera, and M. Karplus. *J. Amer. Chem. Soc.*, 117:11946, 1995.
41. W. L. Jorgensen, D. S. Maxwell, and J. Tiradorives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Amer. Chem. Soc.*, 118:11225–11236, 1996.
42. T. A. Halgren. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comp. Chem.*, 17:490–519, 1996.
43. J. Kong. Combining rules for intermolecular potential paramters. II. Rules for the Lennard-Jones (12-6) potential and the Morse potential. *J. Chem. Phys.*, 59:2464–

2467, 1973.
44. M. Waldman and A. T. Hagler. New combining rules for rare gas van der Waals parameters. *J. Comp. Chem.*, 14:1077, 1993.
45. J. Delhommelle and P. Millié. Inadequacy of the Lorentz-Bertelot combining rules for accurate predictions of equilibrium properties by molecular simulation. *Molec. Phys.*, 99:619–625, 2001.
46. L. Verlet. Computer experiments on classical fluids. I. Thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98, 1967.
47. G. Sutmann and V. Stegailov. Optimization of neighbor list techniques in liquid matter simulations. *J. Mol. Liq.*, 125:197–203, 2006.
48. G. Sutmann and V. Stegailov (to be published).
49. R. W. Hockney. The potential calculation and some applications. *Meth. Comput. Phys.*, 9:136–211, 1970.
50. R. W. Hockney, S. P. Goel, and J. W. Eastwood. Quite high-resolution computer models of a plasma. *J. Comp. Phys..*, 14:148, 1974.
51. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, 64:253, 1921.
52. S. W. de Leeuw, J. M. Perram, and E. R. Smith. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proc. R. Soc. London*, A373:27, 1980.
53. S. W. de Leeuw, J. M. Perram, and E. R. Smith. Simulation of electrostatic systems in periodic boundary conditions. II. Equivalence of boundary conditions. *Proc. R. Soc. London*, A373:57, 1980.
54. T. Darden, D. York, and L. Pedersen. A NlogN method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089, 1993.
55. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577, 1995.
56. R. W. Hockney and J. W. Eastwood. *Computer simulation using particles*. McGraw-Hill, New York, 1981.
57. M. Deserno and C. Holm. How to mesh up Ewald sums. I. a theoretical and numerical comparison of various particle mesh routines. *J. Chem. Phys.*, 109:7678, 1998.
58. M. Deserno and C. Holm. How to mesh up Ewald sums. II. an accurate error estimate for the P3M algorithm. *J. Chem. Phys.*, 109:7694, 1998.
59. L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *J. Comp. Phys.*, 73:325, 1987.
60. H. Cheng, L. Greengard, and V. Rokhlin. A fast adaptive multipole algorithm in three dimensions. *J. Comp. Phys..*, 155:468–498, 1999.
61. C. A. White and M. Head-Gordon. Derivation and efficient implementation of the fast multipole method. *J. Chem. Phys.*, 101:6593–6605, 1994.
62. C. A. White and M. Head-Gordon. Rotating around the quartic angular momentum barrier in fast multipole method calculations. *J. Chem. Phys.*, 105:5061–5067, 1996.
63. C. A. White and M. Head-Gordon. Fractional tiers in fast multipole method calculations. *Chem. Phys. Lett.*, 257:647–650, 1996.
64. H. Dachsel. An improved implementation of the fast multipole method. In *Proceedings of the 4th MATHMOD Vienna*, Vienna, 2003.
65. J. E. Barnes and P. Hut. A hierarchical O($N \log N$) force calculation algorithm.

Nature, 324:446, 1986.
66. S. Pfalzner and P. Gibbon. *Many Body Tree Methods in Physics*. Cambridge University Press, New York, 1996.
67. G. Sutmann and B. Steffen. A particle-particle particle-multigrid method for long-range interactions in molecular simulations. *Comp. Phys. Comm.*, 169:343–346, 2005.
68. G. Sutmann and S. Wädow. A Fast Wavelet Based Evaluation of Coulomb Potentials in Molecular Systems. In U.H.E. Hansmann, editor, *From Computational Biophysics to Systems Biology 2006*, volume 34, pages 185–189, Jülich, 2006. John von Neumann Institute for Computing.
69. N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. Saunders College Publishing, Fort Worth, 1976.
70. R. A. Robinson and R. H. Stokes. *Electrolyte Solutions*. Butterworth, London, 1965.
71. J. W. Perram, H. G. Petersen, and S. W. de Leeuw. An algorithm for the simulation of condensed matter which grows as the 3/2 power of the number of particles. *Molec. Phys.*, 65:875–893, 1988.
72. D. Fincham. Optimisation of the Ewald sum for large systems. *Molec. Sim.*, 13:1–9, 1994.
73. W. Smith. Point multipoles in the Ewald sum. *CCP5 Newsletter*, 46:18–30, 1998.
74. T. M. Nymand and P. Linse. Ewald summation and reaction field methods for potentials with atomic charges, dipoles and polarizabilities. *J. Chem. Phys.*, 112:6152–6160, 2000.
75. G. Salin and J. P. Caillol. Ewald sums for yukawa potentials. *J. Chem. Phys.*, 113:10459–10463, 2000.
76. L. Greengard. *The rapid evaluation of potential fields in particle systems*. MIT press, Cambridge, 1988.
77. L. Greengard. The numerical solution of the N-body problem. *Computers in Physics*, 4:142–152, 1990.
78. L. Greengard. Fast algorithms for classical physics. *Science*, 265:909–914, 1994.
79. J. D. Jackson. *Classical Electrodynamics*. Wiley, New York, 1983.
80. M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover Publ. Inc., New York, 1972.
81. R. K. Beatson and L. Greengard. A short course on fast multipole methods. In M. Ainsworth, J. Levesley, W.A. Light, and M. Marletta, editors, *Wavelets, Multilevel Methods and Elliptic PDEs*, pages 1–37. Oxford University Press, 1997.
82. C. G. Lambert, T. A. Darden, and J. A. Board. A multipole-based algorithm for efficient calculation of forces and potentials in macroscopic periodic assemblies of particles. *J. Comp. Phys..*, 126:274–285, 1996.
83. M. Challacombe, C. A. White, and M. Head-Gordon. Periodic boundary conditions and the fast multipole method. *J. Chem. Phys.*, 107:10131–10140, 1997.
84. S. J. Marrink, E. Lindahl, O. Edholm, and A. E. Mark. Simulation of the spontaneous aggregation of phospholipids into bilayers. *J. Am. Chem. Soc.*, 123:8638, 2001.
85. Michael L. Klein and Wataru Shinoda. Large-scale molecular dynamics simulations of self-assembling systems. *Science*, 321:798 – 800, 2008.
86. A. H. de Vries, A. E. Mark, and S. J. Marrink. Molecular dynamics simulation of the spontaneous formation of a small dppc vesicle in water in atomistic detail. *J. Am.*

Chem. Soc., 126:4488, 2004.
87. G. Srinivas, S. O. Nielsen, P. B. Moore, and M. L. Klein. Molecular dynamics simulations of surfactant self-organization at a solid-liquid interface. *J. Am. Chem. Soc.*, 128:848, 2006.
88. S. J. Marrink X. Periole, Th. Huber and Th. P. Sakmar. Protein-coupled receptors self-assemble in dynamics simulations of model bilayers. *J. Am. Chem. Soc.*, 129:10126, 2007.
89. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.
90. W. L. Jorgensen and J. D. Madura. Temperature and size dependence for Monte Carlo simulations of TIP4P water. *Mol. Phys.*, 56:1381–1392, 1985.
91. H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91:6269, 1987.
92. M. W. Mahoney and W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, 112:8910–8922, 2000.
93. S. J. Marrink, A. H. de Vries, and A. E. Mark. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B*, 108:750, 2004.
94. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P.Tieleman, and A. H. de Vries. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B*, 108:750, 2004.
95. R.L. Henderson. A uniqueness theorem for fluid pair correlation functions. *Phys. Lett.*, 49 A:197–198, 1974.
96. A.P.Lyubartsev and A.Laaksonen. Reconstruction of pair interaction potentials from radial distribution functions. *Comp. Phys. Comm.*, 121-122:57–59, 1999.
97. A.P.Lyubartsev and A.Laaksonen. Determination of pair potentials from ab-initio simulations: Application to liqid water. *Chem. Phys. Lett.*, 325:15–21, 2000.
98. V.Lobaskin, A.P.Lyubartsev, and P.Linse. Effective macroion-macroion potentials in asymmetric electrolytes. *Phys. Rev. E*, 63:020401, 2001.
99. A.P.Lyubartsev and A.Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Comp. Phys. Comm.*, 121-122:57–59, 1999.
100. T.C. Terwilliger. Improving macromolecular atomic models at moderate resolution by automated iterative model building, statistical density modification and refinement. *Acta Cryst.*, D59:1174–1182, 2003.
101. H. F. Trotter. On the product of semi-groups of operators. *Proc. Am. Math. Soc.*, 10:545–551, 1959.
102. O. Buneman. Time-reversible difference procedures. *J. Comp. Phys..*, 1:517–535, 1967.
103. E. Hairer and P. Leone. Order barriers for symplectic multi-value methods. In D. Griffiths, D. Higham, and G. Watson, editors, *Pitman Research Notes in Mathematics*, volume 380, pages 133–149, 1998.
104. D. Okunbor and R. D. Skeel. Explicit canonical methods for Hamiltonian systems. *Math. Comput.*, 59:439–455, 1992.
105. E. Hairer. Backward error analysis of numerical integrators and symplectic methods.

Ann. Numer. Math., 1:107–132, 1994.
106. E. Hairer and C. Lubich. The lifespan of backward error analysis for numerical integrators. *Numer. Math.*, 76:441–462, 1997.
107. S. Reich. Backward error analysis for numerical integrators. *SIAM J. Numer. Anal.*, 36:1549–1570, 1999.
108. D. M. Stoffer. *Some geometrical and numerical methods for perturbed integrable systems*. PhD thesis, Swiss Federal Institute of Technology, Zürich, 1988.
109. J. M. Sanz-Serna M. Calvo. *Numerical Hamiltonian Problems*. Chapman and Hall, London, 1994.
110. R. D. Skeel. Integration schemes for molecular dynamics and related applications. In M. Ainsworth, J. Levesley, and M. Marletta, editors, *The Graduate Student's Guide to Numerical Analysis*, pages 119–176, New York, 1999. Springer.
111. M. E. Tuckerman and W. Langel. Multiple time scale simulation of a flexible model of co_2. *J. Chem. Ohys.*, 100:6368, 1994.
112. P. Procacci, T. Darden, and M. Marchi. A very fast Molecular Dynamics method to simulate biomolecular systems with realistic electrostatic interactions. *J. Phys. Chem.*, 100:10464–10468, 1996.
113. P. Procacci, M. Marchi, and G. L. Martyna. Electrostatic calculations and multiple time scales in molecular dynamics simulation of flexible molecular systems. *J. Chem. Phys.*, 108:8799–8803, 1998.
114. P. Procacci and M. Marchi. Taming the Ewald sum in molecular dynamics simulations of solvated proteins via a multiple time step algorithm. *J. Chem. Phys.*, 104:3003–3012, 1996.
115. J. J. Biesiadecki and R. D. Skeel. Dangers of multiple time step methods. *J. Comp. Phys..*, 109:318–328, 1993.
116. J. L. Scully and J. Hermans. Multiple time steps: limits on the speedup of molecular dynamics simulations of aqueous systems. *Molec. Sim.*, 11:67–77, 1993.
117. B. J. Leimkuhler and R. D. Skeel. Symplectic numerical integrators in constrained Hamiltonian systems. *J. Comp. Phys..*, 112:117–125, 1994.
118. T. Schlick. Some failures and success of long timestep approaches to biomolecular simulations. In P. Deuflhard, J. Hermans, B. J. Leimkuhler, A. Mark, S. Reich, and R. D. Skeel, editors, *Lecture notes in computational science and engineerung. Algorithms for macromolecular modelling*, volume 4, pages 221–250, New York, 1998. Springer.
119. E. Barth and T. Schlick. Overcoming stability limitations in biomolecular danamics. I. Combining force splitting via extrapolation with Langevin dynamics. *J. Chem. Phys.*, 109:1617–1632, 1998.
120. E. Barth and T. Schlick. Extrapolation versus impulse in multiple-timestepping schemes. II. Linear analysis and applications to Newtonian and Langevin dynamics. *J. Chem. Phys.*, 109:1633–1642, 1998.
121. B. Garcia-Archilla, J. M. Sanz-Serna, and R. D. Skeel. Long-time-step methods for oscillatory differential equations. *SIAM J. Sci. Comp.*, 20:930–963, 1998.
122. B. Garcia-Archilla, J. M. Sanz-Serna, and R. D. Skeel. The mollified impulse method for oscillatory differential equations. In D. F. Griffiths and G. A. Watson, editors, *Numerical analysis 1997*, pages 111–123, London, 1998. Pitman.
123. B. Garcia-Archilla, J. M. Sanz-Serna, and R. D. Skeel. The mollified impulse method

for oscillatory differential equations. *SIAM J. Sci. Comp.*, 20:930–963, 1998.
124. J. A. Izaguirre. *Longer time steps for molecular dynamics.* PhD thesis, University of Illinois at Urbana-Champaign, 1999.
125. J. A. Izaguirre, S. Reich, and R. D. Skeel. Longer time steps for molecular dynamics. *J. Chem. Phys.*, 110:9853, 1999.
126. B. V. Chirikov. A universal instability of many-dimensional oscillator systems. *Phys. Rep.*, 52:264–379, 1979.
127. F. Calvo. Largest Lyapunov exponent in molecular systems: Linear molecules and application to nitrogen clusters. *Phys. Rev. E*, 58:5643–5649, 1998.
128. R. F. Warming and B. J. Hyett. The modified equation approach to the stability and accuracy analysis of finite difference methods. *J. Comp. Phys..*, 14:159–179, 1974.
129. M. P. Allen and D. J. Tildesley. *Computer simulation of liquids.* Oxford Science Publications, Oxford, 1987.
130. L. V. Woodcock. Isothermal molecular dynamics calculations for liquid salt. *Chem. Phys. Lett.*, 10:257–261, 1971.
131. W. G. Hoover, A. J. C. Ladd, and B. Moran. High strain rate plastic flow studied via nonequilibrium molecular dynamics. *Phys. Rev. Lett.*, 48:1818–1820, 1982.
132. D. J. Evans, W. G. Hoover, B. H. Failor, B. Moran, and A. J. C. Ladd. Nonequilibrium molecular dynamics via Gauss's principle of least constraint. *Phys. Rev. A*, 28:1016–1021, 1983.
133. D. J. Evans. Computer experiment for nonlinear thermodynamics of Couette flow. *J. Chem. Phys.*, 78:3298–3302, 1983.
134. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684, 1984.
135. H. J. C. Berendsen. Transport properties computed by linear response through weak coupling to a bath. In M. Meyer and V. Pontikis, editors, *Computer Simulation in Materials Science*, pages 139–155, Amsterdam, 1991. Kluwer Academic Publishers.
136. T. Morishita. Fluctuation formulas in molecular dynamics simulations with the weak coupling heat bath. *J. Chem. Phys.*, 113:2976–2982, 2000.
137. T. Schneider and E. Stoll. Molecular dynamics study of a three dimensional one-component model for distortive phase transitions. *Phys. Rev. B*, 17:1302–1322, 1978.
138. H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384, 1980.
139. E. Bonomi. *J. Stat. Phys.*, 39:167, 1985.
140. H. Tanaka, K. Nakanishi, and N. Watanabe. *J. Chem. Phys.*, 78:2626, 1983.
141. M. E. Riley, M. E. Coltrin, and D. J. Diestler. A velocity reset method of simulating thermal motion and damping in gas-solid collisions. *J. Chem. Phys.*, 88:5934–5942, 1988.
142. G. Sutmann and B. Steffen. Correction of finite siye effects in molecular dynamics simulation applied to friction. *Comp. Phys. Comm.*, 147:374–377, 2001.
143. S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511–519, 1984.
144. S. Nosé. A molecular dynamics methood for simulations in the canonical ensemble. *Molec. Phys.*, 52:255–268, 1984.

145. K. Zare and V. Szebehely. Time tranformations for the extended phase space. *Celestial Mech.*, 11:469, 1975.
146. W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.
147. D. J. Evans and G. P. Morris. The isothermal isobaric molecular dynamics ensemble. *Phys. Lett. A*, 98:433–436, 1983.
148. D. J. Evans and G. P. Morris. Isothermal isobaric molecular dynamics. *Chem. Phys.*, 77:63–66, 1983.
149. G. M. Amdahl. Validity of the single-processor approach to achieving large scale computing capabilities. In *AFIPS Conference Proceedings*, volume 30, pages 483–485, Reston, Va., 1967. AFIPS Press.
150. http://trac.mcs.anl.gov/projects/parallel-netcdf.
151. http://trac.mcs.anl.gov/projects/parallel-netcdf/netcdf-api.ps.
152. www.unidata.ucar.edu/packages/netcdf/.
153. http://www.hdfgroup.uiuc.edu/HDF5/projects/archive/WRF-ROMS/Parallel_NetCDF4_Performance.pdf.
154. W. Frings, F. Wolf, and V. Petkov. SIONlib: Scalable parallel I/O for task-local files, 2009. (to be submitted).
155. M. Snir. A note on n-body computations with cutoffs. *Theor. Comput. Syst.*, 37:295–318, 2004.
156. D.E. Shaw. A fast, scalable method for the parallel evaluation of distance limited pairwise particle interactions. *J. Comp. Chem.*, 26:1318–1328, 2005.
157. K.E. Bowers, R.O. Dror, and D.E. Shaw. The midpoint method for parallelization of particle simulations. *J. Chem. Phys.*, 124:184109, 2006.
158. D. Brown, J. H. R. Clarke, M. Okuda, and T. Yamazaki. A domain decomposition parallel processing algorithm for molecular dynamics simulations of polymers. *Comp. Phys. Comm.*, 83:1, 1994.
159. M. Pütz and A. Kolb. Optimization techniques for parallel molecular dynamics using domain decomposition. *Comp. Phys. Comm.*, 113:145–167, 1998.
160. G. Sutmann, R. Winkler, and G. Gompper. Multi-particle collision dynamics coupled to molecular dynamics on massively parallel computers, 2009. (to be submitted).

Monte Carlo and Kinetic Monte Carlo Methods – A Tutorial

Peter Kratzer

Fachbereich Physik and Center for Nanointegration (CeNIDE)
Universität Duisburg-Essen, Lotharstr. 1, 47048 Duisburg, Germany
E-mail: Peter.Kratzer@uni-duisburg-essen.de

1 Introduction

In Computational Materials Science we have learned a lot from molecular dynamics (MD) simulations that allows us to follow the dynamics of molecular processes in great detail. In particular, the combination of MD simulations with density functional theory (DFT) calculations of the electronic structure, as pioneered more than thirty years ago by the work of R. Car and M. Parrinello[1], has brought us a great step further: Since DFT enables us to describe a wide class of chemical bonds with good accuracy, it has become possible to model the microscopic dynamics behind many technological important processes in materials processing or chemistry. It is important to realize that the knowledge we gain by interpreting the outcome of a simulation can be only as reliable as the theory at its basis that solves the quantum-mechanical problem of the system of electrons and nuclei for us. Hence any simulation that aims at predictive power should start from the sub-atomic scale of the electronic many-particle problem. However, for many questions of scientific or technological relevance, the phenomena of interest take place on much larger length and time scales. Moreover, temperature may play a crucial role, for example in phase transitions. Problems of this type have been handled by Statistical Mechanics, and special techniques such as Monte Carlo methods have been developed to be able to tackle with complex many-particle systems[2,3]. However, in the last decade it has been realized that also for those problems that require statistics for a proper treatment, a 'solid basis' is indispensable, i.e. an understanding of the underlying molecular processes, as provided by DFT or quantum-chemical methods. This has raised interest in techniques to combine Monte Carlo methods with a realistic first-principles description of processes in condensed matter.[4]

I'd like to illustrate these general remarks with examples from my own field of research, the theory of epitaxial growth. The term epitaxy means that the crystalline substrate imposes its structure onto some deposited material, which may form a smooth film or many small islands, depending on growth conditions. Clearly, modeling the deposition requires a sample area of at least mesoscopic size, say 1 μm^2, involving tens of thousands of atoms. The time scale one would like to cover by the simulation should be of the same order as the actual time used to deposit one atomic layer, i.e. of the order of seconds. However, the microscopic, atomistic processes that govern the physics and chemistry of deposition, adsorption, and diffusion operate in the length and time domains of 0.1 to 1 nm, and femto-

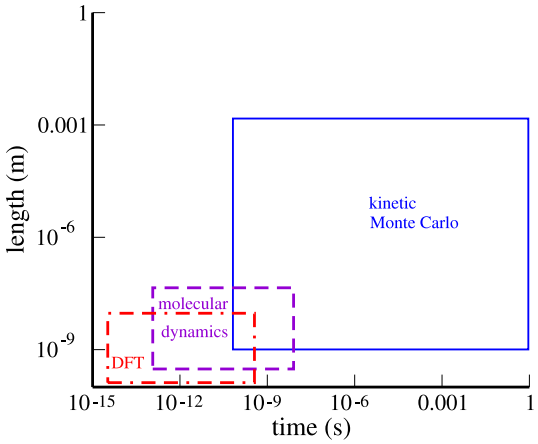


Figure 1. Molecular modeling on the basis of first-principles electronic structure calculations requires to cover the length and time scales from the electronic to the mesoscopic or even macroscopic regime. On the electronic level, density functional theory (DFT) is frequently employed. Molecular dynamics (MD) simulations can be carried out either in combination with DFT, or by using classical forces, which allow one to extend the simulations to bigger length and time scales. The kinetic Monte Carlo method may reach out to very large scales (much depending on the rate constants of the processes relevant to a specific problem), while being able to use input from DFT or MD simulations.

to pico-seconds. Hence incorporating information about atomic processes into modeling of film growth poses the challenge to cover huge length and time scales: from 10^{-10}m to 10^{-6}m and from 10^{-15}s to 10^{0} s (cf. Fig. 1). While smaller length scales, comprising a few hundred atoms, are typically sufficient to gain insight, e.g. about the atomic structure of a step on a surface and its role for chemical reactions and atom diffusion, the gap between the atomic and the practically relevant *time scales* and the crucial role of Statistical Mechanics constitute major obstacles for reliable molecular modeling.

An additional challenge arises due to the complexity of the phenomena to be investigated: One of the fascinating features of epitaxy is the *interplay* of various atomic processes. For example, atoms deposited on an island may be able to overcome the island edge ('step down to the substrate') for specific edge orientations. Thus, while the island changes its shape (and thus the structure of its edges) during growth, this will enable (or disable) material transport between the island top and the substrate, resulting in a transition from two-dimensional to three-dimensional island growth (or vice versa). The possibility that processes may 'trigger' other processes during the evolution of structures can hardly be foreseen or incorporated *a priori* in analytical modeling, but calls for computer simulations using statistical methods.

In epitaxial growth, lattice methods exploiting the two-dimensional periodicity of the substrate lattice are often – but not always – appropriate. Also in other fields of Solid State Physics, mathematical models defined on lattices have been used for a long time. A well-known example is the Ising model in the study of magnetism. It describes the interaction between magnetic moments (spins) sitting on a lattice that can take on two states only ('up' or 'down', represented by variables $s_i = \pm 1$). The Hamiltonian of the Ising model is given

by

$$H(s) = -J_q \sum_i \sum_{j \in n(i)} s_i s_j - \mu_B B \sum_i s_i \tag{1}$$

where $n(i)$ denotes the set of spins interacting with spin i, J_q is the strength of the interaction between spins, q is the number of interacting neighbors ($qJ_q = \text{const} = k_B T_c$, where the last equality is valid in the mean-field approximation), and B is an external magnetic field .

In surface physics and epitaxy, a mathematically equivalent model is used under the name 'lattice Hamiltonian'. It describes fixed sites on a lattice that can be either empty or occupied by a particle (e.g., a deposited atom). The interactions between these particles are assumed to have finite range. The lattice-gas interpretation of the Ising model is obtained by the transformation $s_i = 2c_i - 1, \quad c_i = 0, 1$,

$$H = -4J_q \sum_i \sum_{j \in n(i)} c_i c_j + 2(qJ_q - \mu_B B) \sum_i c_i - N(qJ_q - \mu_B B)\,. \tag{2}$$

For studies of epitaxy, one wishes to describe not only monolayers of atoms, but films that are several atomic layers thick. These layers may not always be complete, and islands and steps may occur on the growing surface. For a close-packed crystal structure, the atoms at a step are chemically less coordinated, i.e., they have fewer partners to form a chemical bond than atoms sitting in flat parts of the surface (= terraces), or atoms in the bulk. Hence, it costs additional energy to create steps, or kinks in the steps. Inspired by these considerations, one can define the so-called solid-on-solid (SOS) model, in which each lattice site is associated with an integer variable, the local surface height h_i. In the SOS model, an energy 'penalty' must be paid whenever two neighbouring lattice sites differ in surface height,

$$H = K_q \sum_i \sum_{j \in n(i)} |h_i - h_j|\,. \tag{3}$$

This reflects the energetic cost of creating steps and kinks. The SOS model allows for a somewhat idealized, but still useful description of the morphology of a growing surface, in which the surface can be described mathematically by a single-valued function h defined on a lattice, i.e., no voids or overhangs in the deposited material are allowed. In the following, we will sometimes refer to one of these three models to illustrate certain features of Monte Carlo simulations. More details about these and other models of epitaxial growth can be found in books emphasizing the statistical-mechanics aspects of epitaxy, e.g. in the textbook by Stanley and Barabasi[5].

In studies of epitaxial growth, model systems defined through a simple Hamiltonian, such as the lattice-gas or the SOS model, have a long history, and numerous phenomena could be described using kinetic Monte Carlo simulations based on these models, dating back to early work by G. H. Gilmer[6], later extended by D. D. Vvedensky[7] and others. For the reader interested in the wealth of structures observed in the evolution of surface morphology, I recommend the book by T. Michely and J. Krug[8]. Despite the rich physics that could be derived from simple models, research in the last decade has revealed that such models are still too narrow a basis for the processes in epitaxial growth. Thanks to more refined experimental techniques, in particular scanning tunneling microscopy[9], but

also thanks to atomistic insights provided by DFT calculations[10,11], we have learned in the last ten years that the processes on the atomic scale are by no means simple. For example, the numerous ways how atoms may attach to an island on the substrate display a stunning complexity. However, kinetic Monte Carlo methods are flexible enough so that the multitude of possible atomic processes can be coded in a simulation program easily, and their macroscopic consequences can be explored.

Apart from simulations of epitaxial growth, thermodynamic as well as kinetic Monte Carlo simulations are a valuable tool in many other areas of computational physics or chemistry. In polymer physics, the ability of Monte Carlo methods to bridge time and length scales makes them very attractive: For example, the scaling properties of polymer dynamics on long time scales (often described by power laws) can be investigated by Monte Carlo simulations.[12] Another important field of applications is in surface chemistry and catalysis[13,14]: Here, Monte Carlo methods come with the bargain that they allow us to study the interplay of a large number of chemical reactions more easily and reliably than the traditional method of rate equations. Moreover, also in this field, feeding information about the individual molecular processes, as obtained e.g. from DFT calculations, into the simulations is a modern trend pursued by a growing number of research groups[15,16].

2 Monte Carlo Methods in Statistical Physics

The term 'Monte Carlo' (MC) is used for a wide variety of methods in theoretical physics, chemistry, and engineering where random numbers play an essential role. Obviously, the name alludes to the famous casino in Monte Carlo, where random numbers are generated by the croupiers (for exclusively non-scientific purposes). In the computational sciences, we generally refer to 'random' numbers generated by a computer, so-called quasi-random numbers [a].

A widely known application of random numbers is the numerical evaluation of integrals in a high-dimensional space. There, the integral is replaced by a sum over function evaluations at discrete support points. These support points are drawn from a random distribution in some compact d-dimensional support $\mathcal{C}$. If the central limit theorem of statistics is applicable, the sum converges, in the statistical sense, towards the value of the integral. The error decreases proportional to the inverse square root of the number of support points, independent of the number of space dimensions. Hence, Monte Carlo integration is an attractive method in particular for integration in high-dimensional spaces.

In Statistical Physics, a central task is the evaluation of the partition function of the canonical ensemble for an interacting system, described by a Hamiltonian H. The contribution of the kinetic energy to H is simple, since it is a sum over single-particle terms. However, calculating the potential energy term $U(x)$ for an interacting many-particle system involves the evaluation of a high-dimensional integral of the type

$$Z = \int_{\mathcal{C}} dx \, \exp\left(-\frac{U(x)}{k_{\mathrm{B}}T}\right) . \tag{4}$$

Here, x stands for a high-dimensional variable specifying the system configuration (e.g., position of all particles). Evaluating this integral by a Monte Carlo method requires special

[a] Concerning the question how a deterministic machine, such as a computer, could possibly generate 'random' numbers, the reader is referred to the numerical mathematics literature, e.g. Ref.[17].

care: Only regions in space where the potential U is small contribute strongly. Hence, using a *uniformly* distributed set of support points would waste a lot of computer resources. Instead, one employs a technique called **importance sampling**. We re-write the partition function

$$Z = \int_{\mathcal{C}} d\mu(x) \tag{5}$$

with the Gibbs measure $d\mu(x) = \exp(-U(x)/(k_{\mathrm{B}}T))\, dx$. The expectation value for an observable is evaluated as the sum over n sampling points in the limit of very dense sampling,

$$\langle O \rangle = \frac{1}{Z} \int_{\mathcal{C}} O(x)\, d\mu(x) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} O(x_i)\mu(x_i)}{\sum_{i=1}^{n} \mu(x_i)}\,. \tag{6}$$

When we generate the n sampling points in configuration space according to their equilibrium distribution, $P_{\mathrm{eq}}(x) = \frac{1}{Z}\exp(-U(x)/(k_{\mathrm{B}}T)) \approx \mu(x_i)/\sum_{i=1}^{n}\mu(x_i)$, we are in position to calculate the thermodynamic average of any observable using

$$\langle O \rangle \approx \frac{1}{n}\sum_{i=1}^{n} O(x_i)\,. \tag{7}$$

The remaining challenge is to generate the support points according to the equilibrium distribution. Instead of giving an explicit description of the equilibrium distribution, it is often easier to think of a stochastic process that tells us how to build up the list of support points for the Gibbs measure. If an algorithm can be judiciously designed in such a way as to retrieve the equilibrium distribution as its limiting distribution, knowing this algorithm (how to add support points) is as good as knowing the final outcome. This 'philosophy' is behind many applications of Monte Carlo methods, both in the realm of quantum physics (Quantum Monte Carlo) and in classical Statistical Physics.

To be more precise, we have to introduce the notion of a **Markov process**. Consider that the system is in a generalized state x_i at some time t_i. (Here, x_i could be a point in a d-dimensional configuration space.) A specific evolution of the system may be characterized by a probability $P_n(x_1, t_1; \ldots; x_n, t_n)$ to visit all the points x_i at times t_i. For example, $P_1(x;t)$ is just the probability of finding the system in configuration x at time t. Moreover, we need to introduce conditional probabilities $p_{1|n}(x_n, t_n|x_{n-1}, t_{n-1}; \ldots; x_1, t_1)$ The significance of these quantities is the probability of finding the system at (x_n, t_n) provided that it has visited already all the space-time coordinates $(x_{n-1}, t_{n-1}) \ldots (x_1, t_1)$. The characteristic feature of a Markov process is the fact that transitions depend on the *previous* step in the chain of events *only*. Hence it is sufficient to consider only *one* conditional probability $p_{1|1}$ for transitions between subsequent points in time. The total probability can then be calculated from the preceeding ones,

$$P_n(x_1, t_1; \ldots; x_n, t_n) = p_{1|1}(x_n, t_n|x_{n-1}, t_{n-1}) P_{n-1}(x_1, t_1; \ldots; x_{n-1}, t_{n-1}) \tag{8}$$

In discrete time, we call such a process a Markov chain. The conditional probabilities of Markov processes obey the **Chapman-Kolmogorov** equation

$$p_{1|1}(x_3, t_3|x_1, t_1) = \int dx_2\, p_{1|1}(x_3, t_3|x_2, t_2) p_{1|1}(x_2, t_2|x_1, t_1) \tag{9}$$

If the Markov process is *stationary*, we can write for its two defining functions

$$\begin{aligned} P_1(x,t) &= P_{\rm eq}(x); \\ p_{1|1}(x_2,t_2|x_1,t_1) &= p_t(x_2|x_1); \qquad t = t_2 - t_1\,. \end{aligned}$$

Here $P_{\rm eq}$ is the distribution in thermal equilibrium, and p_t denotes the transition probability within the time interval t from a state x_1 to a state x_2.

Using the Chapman-Kolmogorov equation for p_t, we get

$$p_{t+t_0}(x_3|x_1) = \int dx_2\, p_{t_0}(x_3|x_2) p_t(x_2|x_1)\,. \tag{10}$$

When we consider a discrete probability space for x_i, the time evolution of the probability proceeds by matrix multiplication, the p_t being matrices transforming one discrete state into another. We now want to derive the differential form of the Chapman-Kolmogorov equation for stationary Markov processes. Therefore we consider the case of small time intervals t_0 and write the transition probability in the following way,

$$p_{t_0}(x_3|x_2) \approx (1 - w_{\rm tot}(x_2)t_0)\delta(x_3 - x_2) + t_0 w(x_3|x_2) + \ldots, \tag{11}$$

up to to terms that vanish faster than linear in t_0. This equation defines $w(x_3|x_2)$ as the transition rate (transition probability per unit time) to go from x_2 to x_3. In the first term, the factor $(1 - w_{\rm tot}(x_2)t_0)$ signifies the probability to remain in state x_2 up to time t_0. That means that $w_{\rm tot}(x_2)$ is the total probability to leave the state x_2, defined as

$$w_{\rm tot}(x_2) = \int dx_3\, w(x_3|x_2). \tag{12}$$

Inserting this into the Chapman-Kolmogorov equation results in

$$p_{t+t_0}(x_3|x_1) = (1 - w_{\rm tot}(x_3)t_0) p_t(x_3|x_1) + t_0 \int dx_2\, w(x_3|x_2) p_t(x_2|x_1); \tag{13}$$

and hence we obtain

$$\frac{p_{t+t_0}(x_3|x_1) - p_t(x_3|x_1)}{t_0} = \int dx_2\, w(x_3|x_2) p_t(x_2|x_1) - \int dx_2 w(x_2|x_3) p_t(x_3|x_1), \tag{14}$$

in which we have used the definition of $w_{\rm tot}$. In the limit $t_0 \to 0$ we arrive at the **master equation**, that is the differential version of the Chapman-Kolmogorov equation,

$$\frac{\partial}{\partial t} p_t(x_3|x_1) = \int dx_2\, w(x_3|x_2) p_t(x_2|x_1) - \int dx_2 w(x_2|x_3) p_t(x_3|x_1)\,. \tag{15}$$

It is an integro-differential equation for the transition probabilities of a stationary Markov process. In the following we do not assume stationarity and choose a $P_1(x_1,t) \neq P_{\rm eq}(x)$, but keep the assumption of time-homogeneity of the transition probabilities, i.e., it is assumed that they only depend on time differences. Then, we can multiply this equation by $P_1(x_1,t)$ and integrate over x_1 to get a master equation for the probability density itself:

$$\frac{\partial}{\partial t} P_1(x_3,t) = \int dx_2\, w(x_3|x_2) P_1(x_2,t) - \int dx_2 w(x_2|x_3) P_1(x_3,t) \tag{16}$$

One way to fulfill this equation is to require **detailed balance**, i.e., the net probability flux between every pair of states in equilibrium is zero,

$$\frac{w(x|x')}{w(x'|x)} = \frac{P_{\rm eq}(x)}{P_{\rm eq}(x')} \,. \tag{17}$$

For thermodynamic averages in the canonical ensemble we have $P_{\rm eq}(x) = \frac{1}{Z}\exp(-H(x)/(k_{\rm B}T))$, and hence

$$\frac{w(x|x')}{w(x'|x)} = \exp\left(-(H(x) - H(x'))/(k_{\rm B}T)\right) \,. \tag{18}$$

When we use transition probabilities in our Monte Carlo simulation that fulfill detailed balance with the desired equilibrium distribution, we are sure to have

$$P_1(x, t \to \infty) = P_{\rm eq}(x) \,. \tag{19}$$

Since the detailed balance condition can be fulfilled in many ways, the choice of transition rates is therefore not unique. Common choices for these rates are

- the **Metropolis rate**

  $w(x'|x) = w_0(x'|x)\min([1; \exp\left(-(H(x') - H(x))/(k_{\rm B}T)\right)]$

- the **Glauber rate**

  $w(x'|x) = w_0(x'|x)\frac{1}{2}\left\{1 - \tanh\left[\exp\left(-(H(x') - H(x))/(k_{\rm B}T)\right)\right]\right\}$

Both choices obey the detailed balance condition. With either choice, we still have the freedom to select a factor $w_0(x'|x) = w_0(x|x')$. This can be interpreted as the probability to choose a pair of states $x,\ x'$ which are connected through the specified move. In an Ising model simulation, each state x corresponds to one particular arrangement of all spins on all the lattice sites. The states x and x' may, for instance, just differ in the spin orientation on one lattice site. Then, the freedom in $w_0(x'|x)$ corresponds to the freedom to select any single spin (with a probability of our choice), and then to flip it (or not to flip it) according to the prescription of the rate law.

Let's illustrate the general considerations by an example. Suppose we want to calculate the magnetization of an Ising spin model at a given temperature. Hence we have to simulate the canonical ensemble using the **Monte Carlo algorithm**. The steps are:

- generate a starting configuration s_0,
- select a spin, s_i, at random,
- calculate the energy change upon spin reversal ΔH,
- calculate the probability $w(\uparrow, \downarrow)$ for this spin-flip to happen, using the chosen form of transition probability (Metropolis or Glauber),
- generate a uniformly distributed random number, $0 < \rho < 1$; if $w > \rho$, flip the spin, otherwise retain the old configuration.

When the Metropolis rate law has been chosen, proposed spin flips are either accepted with probability w, or discarded with probability $1-w$. After some transient time, the system will come close to thermodynamic equilibrium. Then we can start to record time averages of some observable O we are interested in, e.g., the magnetization. Due to the in-built properties of the rate law, this time average will converge, in the statistical sense, to the thermodynamic ensemble average $\langle O \rangle$ of the observable O.

The prescription for the Monte Carlo method given so far applies to non-conserved observables (e.g., the magnetization). For a conserved quantity (e.g., the concentration of particles), one uses **Kawasaki dynamics**:

- choose a pair of (neighboring) spins [b]
- exchange the spins subject to the Metropolis acceptance criterion

Since this algorithm guarantees particle number conservation, it recommends itself for the lattice-gas interpretation of the Ising model. In simulations of epitaxial growth, one may work with either conserved or non-conserved particle number, the latter case mimicking desorption or adsorption events of particles.

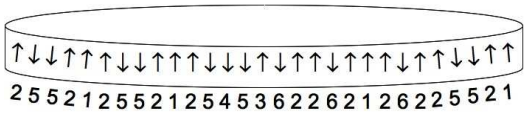


Figure 2. Illustration of the N-fold way algorithm for a one-dimensional Ising chain of spins. A particular configuration for one moment in time is shown. The local environments of all spins fall in one of six classes, indicated by the numbers. Periodic boundary conditions are assumed.

Table 1. Classification on spins in a 6-fold way for a periodic Ising chain. The leftmost column gives the number n_i of spins in each class for the particular configuration shown in Fig. 2. The rates can be normalised to unity by setting $w_0 = \{[n_1 \exp(-2\mu_B/(k_\mathrm{B}T)) + n_4 \exp(2\mu_B/(k_\mathrm{B}T))] \exp(-4J_q/(k_\mathrm{B}T)) + n_2 \exp(-2\mu_B/(k_\mathrm{B}T)) + n_5 \exp(2\mu_B/(k_\mathrm{B}T)) + [n_3 \exp(-2\mu_B/(k_\mathrm{B}T)) + n_6 \exp(2\mu_B/(k_\mathrm{B}T))] \exp(4J_q/(k_\mathrm{B}T))\}^{-1}$.

class	central spin	neighbors	rate w_i	class members n_i
1	$\uparrow$	$\uparrow,\uparrow$	$w_0 \exp\big(-(4J_q + 2\mu_B B)/(k_\mathrm{B}T)\big)$	4
2	$\uparrow$	$\uparrow,\downarrow$	$w_0 \exp\big(-2\mu_B B/(k_\mathrm{B}T)\big)$	12
3	$\uparrow$	$\downarrow,\downarrow$	$w_0 \exp\big((4J_q - 2\mu_B B)/(k_\mathrm{B}T)\big)$	1
4	$\downarrow$	$\downarrow,\downarrow$	$w_0 \exp\big(-(4J_q - 2\mu_B B)/(k_\mathrm{B}T)\big)$	1
5	$\downarrow$	$\uparrow,\downarrow$	$w_0 \exp\big(2\mu_B B/(k_\mathrm{B}T)\big)$	8
6	$\downarrow$	$\uparrow,\uparrow$	$w_0 \exp\big((4J_q + 2\mu_B B)/(k_\mathrm{B}T)\big)$	3

[b] This means $w_0(s'|s) = 1/(2dn)$, i.e. we first choose a spin at random, and then a neighbor on a d-dimensional simple cubic lattice with n sites at random.

3 From MC to kMC: The N-Fold Way

The step forward from the Metropolis algorithm to the algorithms used for kinetic Monte Carlo (kMC) simulations originally resulted from a proposed speed-up of MC simulations: In the Metropolis algorithm, trial steps are sometimes discarded, in particular if the temperature is low compared to typical interaction energies. For this case, Bortz, Kalos and Lebowitz suggested in 1975 the N-fold way algorithm[18] that avoids discarded attempts. The basic idea is the following: In an Ising model or similar models, the interaction energy, and thus the rate of spin-flipping, only depends on the nearest-neighbor configuration of each spin. Since the interaction is short-ranged, there is only a small number of local environments (here: spin triples), each with a certain rate w_i for flipping the 'central' spin of the triple. For example, in one dimension, i.e. in an Ising chain, an 'up' spin may have both neighbors pointing 'up' as well, or both neighbors pointing 'down', or alternate neighbors, one 'up', one 'down'. An analogous classification holds if the selected (central) spin points 'down'. All local environments fall into one of these six classes, and there are six 'types' of spin flipping with six different rate constants. For a given configuration of a spin chain, one can enumerate how frequently each class of environment occurs, say, n_i times, $i = 1, \ldots 6$. This is illustrated in Table 1 for the configuration shown in Fig. 2. Now the **N-fold way algorithm** works like this:

1. first select a class i with a probability given by $n_i w_i / \sum_i w_i n_i$ using a random number ρ_1;
2. then, select one process (i.e., one spin to be flipped) of process type i, choosing with equal probability among the representatives of that class, by using another random number ρ_2;
3. execute the process, i.e. flip the spin;
4. update the list of n_i according to the new configuration.

The algorithm cycles through this loop many times, without having to discard any trials, thereby reaching thermal equilibrium in the spin system. The prescription can easily be generalized to more dimensions; e.g., to a two-dimensional square lattice, where we have ten process types.[c]

To go all the way from MC to kMC, what is still missing is the aspect of *temporal evolution*. In a MC simulation, we may count the simulation steps. However, the foundation of the method lies in *equilibrium* statistical physics. Once equilibrium is reached, time has no physical meaning. Therefore no physical basis exists for identifying simulation steps with physical time steps in the conventional Monte Carlo methods. In order to address *kinetics*, i.e. to make a statement how fast a system reaches equilibrium, we need to go beyond that, and take into account for the role of time. Here, some basic remarks are in place. In order to be able to interpret the outcome of our simulations, we have to refer to some assumptions about the *separation of time scales*: The shortest time scale in the problem is given by the time it takes for an elementary process (e.g., a spin flip) to proceed. This time scale should be clearly separated from the time interval between two processes

[c] Each 'central' spin has four neighbors, and the number of neighbors aligned with the 'central' spin may vary between 0 and 4. Taking into account that the central spin could be up or down, we end up with ten process types.

taking place at the same spin, or within the local environment of one spin. This second time scale is called the waiting time between two subsequent events. If the condition of time scale separation is not met, the remaining alternative is a simulation using (possibly accelerated) molecular dynamics (see Section 4.2). If time scale separation applies, one of the basic requirements for carrying out a kMC simulation is fulfilled. The advantage is that kMC simulations can be run to simulate much longer physical time intervals at even lower computational cost than molecular dynamics simulations (see Fig. 1). Moreover, one can show that the waiting time, under quite general assumptions, follows a Poissonian distribution[19]. For the Ising chain, each process type has a different waiting time $\tau_i = w_i^{-1}$ that is proportional to some power of $\exp(J/(k_{\mathrm{B}}T))$. For other applications of interest, the waiting times of various process types may be vastly different. In epitaxial growth, for instance, the time scale between two adsorption or desorption events is usually much longer than the time scale for surface diffusion between two adjacent sites. For macromolecules in the condensed phase, vibrational motion of a molecular side group may be fast, while a rotation of the whole molecule in a densely packed environment may be very slow, due to steric hindrance. In a kinetic simulation, we would like to take all these aspects into account. We need a simulation method that allows us to **bridge time scales** over several orders of magnitude.

Following the N-fold way for the Ising model, it is easy to calculate the *total* rate R, i.e., the probability that some event will occur in the whole system per unit time. It is the sum of all rates of individual processes, $R = \sum_i n_i w_i$. The average waiting time between any two events occurring in the system as a whole is given by R^{-1}. This allows us to associate a time step of (on average) $\Delta t = R^{-1}$ with one step in the simulation. Note that the actual length of this time step may change (and in general does so) during the simulation, since the total rate of all processes accessible in a certain stage of the simulation may change. Therefore, this variant of the kMC method is sometimes also called the 'variable step size' method in the literature. More realistically, the time step Δt should not be identified with its average value, but should should be drawn from a Poissonian distribution. This is practically realised by using the expression $\Delta t = -R^{-1} \log \rho_3$ with a random number $0 < \rho_3 < 1$. For a two-dimensional problem (e.g., a lattice-gas Hamiltonian), the N-fold way algorithm is explained in the flowchart of Fig. 4.

The distinction between MC and kMC simulations is best understood by considering the following points: In kMC, the possible configurations of the system, i.e. the micro-states contributing to the macro-state of a statistical ensemble, need to be enumerable, in order to be able to build up a list of process types, as in Table 1. In a MC simulation, on the other hand, there is no limit on the number of micro-states – they even need not be known to start the simulation. For this reason, the MC algorithm can be applied to problems with a huge configuration space, e.g. to protein folding, where a kMC simulation would not be feasible. In advantage over MC, a kMC simulation allows us to assign the meaning of a physical time to the simulation steps. Of course, in order to make use of this advantage, we need to provide as input the rates of all relevant individual processes. Obtaining information about all these rates is a difficult task; this is why kMC simulations are less common than MC simulations. The best way for getting values for the individual rates is by performing molecular dynamics simulations, possibly with first-principles electronic structure methods such as DFT. This will be explained in more detail in Section 4.2.

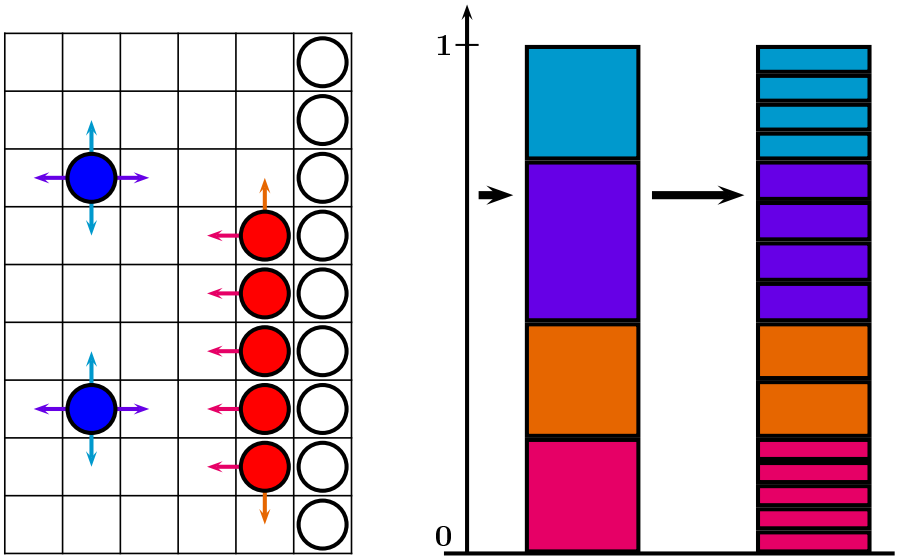


Figure 3. Principle of process-type list algorithm. There are certain types of processes, indicated by colors in the figure: diffusion on the substrate, diffusion along a step, detachment from a step, ...(left scheme). Each type is described by its specific rate constant, but processes of the same type have the same rate constant. Hence, the list of all processes can be built up as a nested sum, first summing over processes of a given type, then over the various types. The selection of a process by a random number generator (right scheme) is realised in two steps, as indicated by the thick horizontal arrows, where the second one selects among equal probabilities.

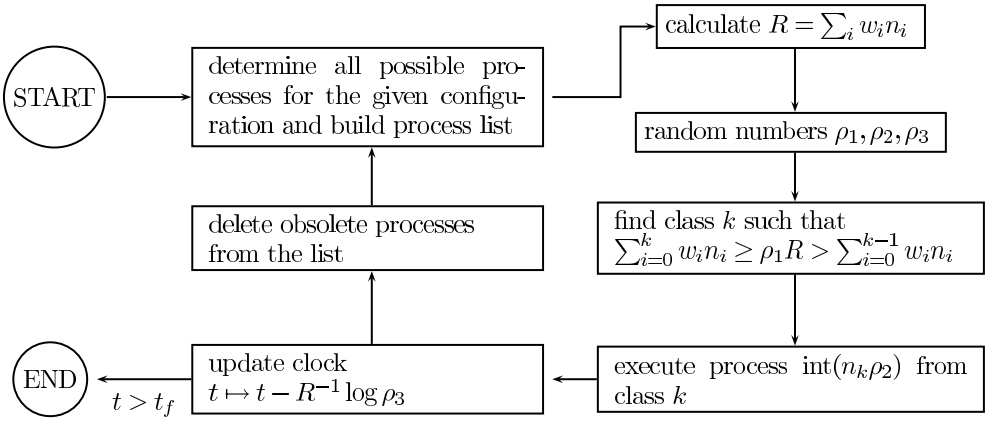


Figure 4. Flow chart for the process-type list algorithm.

Finally, we note that a kMC simulation provides a particular solution of the master equation in a stochastic sense; by averaging over many kMC runs we obtain probabilities (for the system being in a specific micro-state) that evolve in time according to Eq. (15).

3.1 Algorithms for kMC

In the kMC algorithm outlined above, the process list is ordered according to the process type; therefore I refer to it as the **process-type list** algorithm. In the practical implementation of kMC algorithms, there are two main concerns that affect the computational efficiency: First, the selection of a suitable algorithm depends on the question *how many* process types are typically active at each moment in time. The second concern is to find an efficient scheme of representing and updating the data. For an efficient simulation, it is essential to realise that the updating of the process list, step 4 of the algorithm described in the previous Section, only modifies those entries that have changed due to the preceding simulation step. A complete re-build of the list after each simulation step would be too time-consuming. As the interactions are short-ranged, a local representation of the partial rates associated with lattice sites is most efficient. On the other hand, the process-type list groups together processes having the same local environment, disregarding where the representatives of this class of processes (the spins or atoms) are actually located on the lattice. Hence, updating the process list requires replacement of various entries that originate from a small spatial region, but are scattered throughout the process list. To handle this task, a subtle system of referencing the entries is required in the code. This is best realised in a computer language such as C by means of pointers. An example of a kMC code treating the SOS model is available from the author upon request.

The two-step selection of the next event, as illustrated in Fig. 3, makes the process-type list advantageous for simulation with a moderate number (say, $N < 100$) of process types. This situation is encountered in many simulations of epitaxial crystal growth using an SOS model[6], but the process-list algorithm also works well for more refined models of crystal growth[20,21]. In the first selection step, we need to compare the random number ρ_1 to at most N partial sums, namely the expressions $\sum_i^k n_i w_i$ for $k = 1, \ldots N$. The second selection step chooses among equally probable alternatives, and requires no further comparing of numbers. Thus, the total number of numerical comparisons needed for the selection is at most N, assuring that this selection scheme is computationally efficient.

In some applications, the kMC algorithm needs to cope with a vast number of different process types. For example, such a situation is encountered in epitaxy when the interaction is fairly long-ranged[22], or when rates depend on a continuous variable, such as the local strain in an elastically deformed lattice. Having to choose among a huge number of process types makes the selection based on the process-type list inefficient. Instead, one may prefer to work directly with a local data representation, and to do the selection of a process in real space. One may construct a suitable multi-step selection scheme by grouping the processes in real space, as suggested by Maksym[23]. Then, one will first draw a random number ρ_1 to select a region in space, then use a second random number ρ_2 to select a particular processes that may take place in this region. Obviously, such a selection scheme is independent of the number of process types, and hence can work efficiently even if a huge number of process types is accessible. Moreover, it can be generalized further: It is always possible to select one event out of $N = 2^k$ possibilities by making k alternative decisions. This comes with the additional effort of having to draw k random numbers $\rho_i, i = 1, \ldots k$, but has the advantage that one needs to compare to $k = \log_2 N$ partial sums only. The most efficient way of doing the selection is to arrange the partial sums of individual rates on a **binary tree**. This allows for a fast hierarchical update of the partial sums associated with each branch point of the tree after a process has been executed.

Finally, I'd like to introduce a third possible algorithm for kMC simulations that abandons the idea of the N-fold way. Instead, it emphasizes the aspect that each individual event, in as far as it is independent from the others, occurs after a random waiting time according to a Poissonian distribution. I refer to that algorithm as the **time-ordered list** algorithm, but frequently it is also called the 'first reaction' method[24,25]. It proceeds as follows:

1. At time t, assign a prospective execution time $t + t_i$ to each individual event, drawing the random waiting time t_i from a Poissonian distribution;

2. sort all processes according to prospective execution time (This requires only $\log_2 N$ comparisons, if done in a 'binary tree');

3. always select the *first* process of the time-ordered list and execute it;

4. advance the clock to the execution time, and update the list according to the new configuration.

This algorithm requires the t_i to be Poissonian random numbers, i.e. to be distributed between 0 and ∞ according to an exponentially decaying distribution function. Hence it may be advisable to use a specially designed random number generator that yields such a distribution. The time-ordered-list algorithm differs from the two others in the fact that the selection step is deterministic, as always the top entry is selected. Yet, its results are completely equivalent to the two other algorithms, provided the common assumption of Poissonian processes holds: In a Poissonian processes, the waiting times are distributed exponentially.[19] In the time-ordered list algorithm, this is warranted explicitly for each event by assigning its time of execution in advance. In the other algorithms, the clock, i.e., the 'global' time for all events, advances according to a Poissonian process. The individual events are picked at random from a list; however, it is known from probability theory that drawing a low-probability event from a long list results in a Poissonian distribution of the time until this event gets selected. Hence, not only the global time variable, but also the waiting time for an individual event follows a Poissonian distribution, as it should be.

The time-ordered list algorithm appears to be the most general and straightforward of the three algorithms discussed here. But again, careful coding is required: As for the process-type list, updating the time-ordered list requires deletion or insertion of entries scattered all over the list. Suggestions how this can be achieved, together with a useful discussion of algorithmic efficiency and some more variants of kMC algorithms can be found in Ref. [24].

In principle, kMC is an inherently serial algorithm, since in one cycle of the loop only one process can be executed, no matter how large the simulation area is. Nonetheless, there have been a number of attempts to design **parallel kMC algorithms**, with mixed success. All these parallel versions are based on a partitioning, in one way or another, of the total simulation area among parallel processors. However, the existence of a global 'clock' in the kMC algorithm would prevent the parallel processors from working independently. In practice, most parallel kMC algorithms let each processor run independently for some time interval small on the scale of the whole simulation, but still long enough to comprise of a large number of events. After each time interval the processors are synchronised and exchange data about the actual configurations of their neighbours. Typically,

this communication among processors must be done very frequently during program execution. Hence the parallel efficiency strongly depends on latency and bandwidth of the communication network. There are a number of problems to overcome in parallel kMC: Like in any parallel simulation of discrete events, the 'time horizon' of the processors may proceed quite inhomogeneously, and processors with little work to do may wait a long time until other, more busy processors have caught up. Even a bigger problem may arise from events near the boundary of processors: Such events may turn out to be impossible after the synchronisation has been done, because the neighbour processor may have modified the boundary region prior to the execution of the event in question. Knowing the actual state of the neighbouring processor, the event should have occurred with a different rate, or maybe not at all. In this case, a 'roll-back' is required, i.e., the simulation must be set back to the last valid event before the conflicting boundary event occurred, and the later simulation steps must be discarded. While such roll-backs are manageable in principle, they may lead to a dramatic decrease in the efficiency of a parallel kMC algorithm. Yet, one may hope that the problems can be kept under control by choosing a suitable synchronisation interval. This is essentially the idea behind the 'optimistic' synchronous relaxation algorithm[26,27]. Several variants have been suggested that sacrifice less efficiency, but pay the price of a somewhat sloppy treatment of the basic simulation rules. In the semi-rigorous synchronous sublattice algorithm[28], the first, coarse partitioning of the simulation area is further divided into sublattices, e.g. like the black and white fields on the checkerboard. Then, in each time interval between synchronisations, events are alternately simulated *only* within one of the sublattices ('black or 'white'). This introduces an arbitrary rule additionally restricting the possible processes, and thus may compromise the validity of the results obtained, but it allows one to minimise or even completely eliminate conflicting boundary events. Consequently, 'roll backs' that are detrimental to the parallel efficiency can be reduced or avoided. However, even when playing such tricks, the scalability of parallel kMC simulations for typical tasks is practically limited to four or eight parallel processors with the currently available parallel algorithms.

4 From Molecular Dynamics to kMC: The Bottom-up Approach

So far, we have been considering model systems. In order to make the formalism developed so far useful for chemistry or materials science, we need to describe how the relevant processes and their rate constants can be determined in a sensible way for a certain system or material. This implies bridging between the level of a molecular dynamics description, where the system is described by the positions and momenta of all particles, and the level of symbolic dynamics characteristic of kMC. For a completely general case, this may be a daunting task. For the special case of surface diffusion and epitaxial growth, it is typically a complex, but manageable problem. On the atomistic level, the motion of an adatom on a substrate is governed by the potential-energy surface (PES), which is the potential energy experienced by the diffusing adatom

$$E^{\mathrm{PES}}(X_{\mathrm{ad}}, Y_{\mathrm{ad}}) = \min_{Z_{\mathrm{ad}}, \{\mathbf{R}_I\}} U(X_{\mathrm{ad}}, Y_{\mathrm{ad}}, Z_{\mathrm{ad}}, \{\mathbf{R}_I\})\,. \tag{20}$$

Here $U(X_{\mathrm{ad}}, Y_{\mathrm{ad}}, Z_{\mathrm{ad}}, \{\mathbf{R}_I\})$ is the potential energy of the atomic configuration specified by the coordinates $(X_{\mathrm{ad}}, Y_{\mathrm{ad}}, Z_{\mathrm{ad}}, \{\mathbf{R}_I\})$. According to Eq. (20), the PES is the minimum

of the potential energy with respect to both the adsorption height, denoted by $Z_{\rm ad}$, and all coordinates of the substrate atoms, denoted by $\{\mathbf{R}_I\}$. The potential energy U can in principle be calculated from any theory of the underlying microscopic physics. Presently, calculations of the electronic structure using density functional theory (DFT) are the most practical means of obtaining an accurate PES. Within DFT, the energy U in Eq. (20) is referred to as the total energy (of the combined system of electrons and nuclei). The above expression is valid at zero temperature. At realistic temperatures, the free energy should be considered instead of U. If we assume for the moment that the vibrational contributions to the free energy do not change the topology of the PES significantly, the minima of the PES represent the stable and metastable sites of the adatom.

Next, we need to distinguish between crystalline solids on the one hand, and amorphous solids or liquids on the other hand. For a crystalline substrate, one will frequently (but not always) encounter the situation that the minima of the PES can be mapped in some way onto (possibly a subset of) lattice sites. The lattice sites may fall into several different classes, but it is crucial that all lattice sites belonging to one class are always connected in the same way to neighbouring sites. Then the dynamics of the system can be considered as a sequence of discrete transitions, starting and ending at lattice sites (lattice approximation). The sites belonging to one class all have the same number of connections, and each connection, i.e. each possible transition, is associated with a rate constant. Methods for amorphous materials going beyond this framework will be discussed later in Section 4.2.

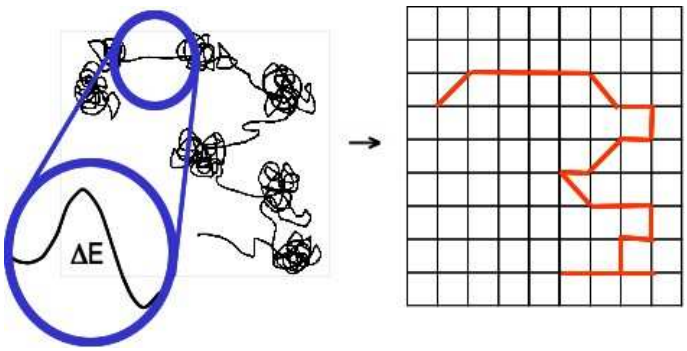


Figure 5. Mapping of the diffusive Brownian motion of a particle on a substrate onto hopping between lattice sites. The particle's trajectory spends most of its time near the minima. In the blow-up of the small piece of the trajectory that crosses a saddle point between two minima, the energy profile along the reaction path is shown. Along the path, the saddle point appears as a maximum of the energy associated with the energy barrier ΔE that must be overcome by the hopping particle.

In surface diffusion, a widely used approximation for calculating the rate constants for the transitions between lattice sites is the so-called Transition State Theory (TST). As this is the 'workhorse' of the field, we will describe it first. The more refined techniques presented later can be divided into two classes: techniques allowing for additional complexity but building on TST for the individual rates, and attempts to go beyond TST in the evaluation of rate constants.

4.1 Determining rate constants from microscopic physics

In order to go from a microscopic description (typical of a molecular dynamics simulation) to a meso- or macroscopic description by kinetic theories, we start by dividing the phase space of the system into a 'more important' and a 'less important' part. In the important part, we'll persist to follow the motion of individual degrees of freedom. One such degree of freedom is the so-called 'reaction coordinate' that connects the initial state x with a particular final state x' (both minima of the PES) we are interested in. The reaction coordinate may be a single atomic coordinate, a linear combination of atomic degrees of freedom, or, most generally, even a curved path in configuration space. The degrees of freedom in the 'less important' part of the system are no more considered individually, but lumped together in a 'heat bath', a thermal ensemble characterised by a temperature. In surface diffusion, the mentioned division of the system into 'more important' and 'less important' parts could (but need not) coincide with the above distinction between the coordinates of the adsorbate atom, $(X_{\rm ad}, Y_{\rm ad}, Z_{\rm ad})$, and the substrate atoms, $\{\mathbf{R}_I\}$. Here, as in most cases, the distinction between the two parts is not unique; there is some arbitrariness, but retaining a sufficiently large part whose dynamics is treated explicitly should yield results that are independent of the exact way the division is made. Of course, the two parts of the system are still coupled: The motion along the reaction path may dissipate energy to the heat bath, an effect that is usually described by a friction constant λ. Likewise, thermal fluctuations in the heat bath give rise to a fluctuating force acting on the reaction coordinate. [d]

Now we want to calculate the rate for the system to pass from the initial to the final state, at a given temperature of the heat bath. For the case of interest, the two states are separated by an energy barrier (or, at least, by a barrier in the free energy). For this reason, the average waiting time for the transition is much longer than typical microscopic time scales, e.g. the period of vibrational motion of a particle in a minimum of the potential. In other words, the transition is an infrequent event; and trying to observe it in a molecular dynamics simulation that treats the whole system would be extremely cumbersome. Therefore, we turn to rate theory to treat such rare events. Within the setting outlined so far, there is still room for markedly different behaviour of different systems, depending on the coupling between the system and the heat bath, expressed by the friction constant λ, being 'strong' or 'weak'. For a general discussion, the reader is referred to a review article[29]. In the simplest case, if the value of the friction constant is within some upper and lower bounds, one can show that the result for the rate is independent of the value of λ. This is the regime where Transition State Theory is valid[30,31]. If the condition is met, one can derive a form of the rate law

$$w_i = \frac{k_{\rm B}T}{h} \exp(-\Delta F_i/(k_{\rm B}T)) \,, \tag{21}$$

for the rate w_i of a molecular process i. Here, i is a shorthand notation for the pair of states x, x', i.e., $w_i \equiv w(x'|x)$. In this expression, ΔF_i is the difference in the free energy between the maximum (saddle point) and the minimum (initial geometry) of the potential-energy surface along the reaction path of the process i. T is the temperature, $k_{\rm B}$

[d] The friction force and the fluctuations of the random thermal force are interrelated, as required by the fluctuation-dissipation theorem.

the Boltzmann constant, and h the Planck constant. Somewhat oversimplifying, one can understand the expression for the rate as consisting of two factors: The first factor describes the inverse of the time it takes for a particle with thermal velocity to traverse the barrier region. The second factor accounts for the (small) probability that a sufficient amount of energy to overcome the barrier is present in the reaction coordinate, i.e., various portions of energy usually scattered among the many degrees of freedom of the heat bath happen by chance to be collected for a short moment in the motion along the reaction coordinate. The probability for this (rather unlikely) distribution of the energy can be described by the Boltzmann factor in Eq. (21). The assumptions for the applicability of TST imply that the free energy barrier ΔF_i should be considerably higher than $k_{\mathrm{B}}T$. Speaking in a sloppy way, one could say that for such high barriers 'gathering the energy to overcome the barrier' and 'crossing the barrier' are two independent things, reflected in the two factors in the TST expression of the rate.

The free energy of activation ΔF_i needed by the system to move from the initial position to the saddle point may be expressed in two ways: Using the fact that the free energy is the thermodynamic potential of the canonical ensemble, ΔF_i may be expressed by the ratio of partition functions,

$$\Delta F_i = k_{\mathrm{B}} \log \left(\frac{Z_i^{(0)}}{Z_i^{\mathrm{TS}}} \right) . \tag{22}$$

where $Z_i^{(0)}$ is the partition function for the m 'important' degrees of freedom of the system in its initial state, and Z_i^{TS} is the partition function for a system with $m-1$ degrees of freedom located at the transition state (saddle point). This partition function must be evaluated with the constraint that only motion in the hyperplane perpendicular to the reaction coordinate are permitted; hence the number of degrees of freedom is reduced by one.

Alternatively, one may use the decomposition

$$\Delta F_i = \Delta E_i - T \Delta S_i^{\mathrm{vib}} . \tag{23}$$

Here ΔE_i is the difference of the internal energy (the (static) total energy and the vibrational energy) of the system at the saddle point and at the minimum, and $\Delta S_i^{\mathrm{vib}}$ is the analogous difference in the vibrational entropy. The rate of the process i can be cast as follows,

$$w_i = w_i^{(0)} \exp(-\Delta E_i / k_{\mathrm{B}} T) , \tag{24}$$

where $w_i^{(0)} = (k_{\mathrm{B}}T/h) \exp(\Delta S_i^{\mathrm{vib}}/k_{\mathrm{B}})$ is called the attempt frequency.

The two basic quantities in Eq. (24), $w_i^{(0)}$ and ΔE_i, can both be obtained from DFT calculations. If we restrict ourselves to single-particle diffusion and neglect the contribution of thermal vibrational energy, ΔE_i can be read off directly from the PES. To obtain the value of the attempt frequency, one may perform molecular dynamics simulations of the canonical ensemble, sampling the partition functions $Z_i^{(0)}$ and Z_i^{TS}. For a computationally simpler, but less accurate approach, one may expand the PES in a quadratic form around the minimum and the saddle point. In this approximation, the partition functions in Eq. (22), which then equal those of harmonic oscillators, may be evaluated analytically,

and one arrives at the frequently used expression

$$w_i^{(0)} = \frac{\prod_{k=1}^{n} \omega_{k,i}^{(0)}}{\prod_{k=1}^{n-1} \omega_{k,i}^{\mathrm{TS}}} \,. \tag{25}$$

Here $\omega_{k,i}^{(0)}$ and $\omega_{k,i}^{\mathrm{TS}}$ are the frequencies of the normal modes at the initial minimum and at the transition state of process i, respectively. Note that the attempt frequency, within the harmonic approximation, is independent of temperature.

Finally, we will briefly comment on the validity of TST for processes in epitaxial growth. For surface diffusion of single adatoms, it has been shown for the case of Cu/Cu(111) that TST with thermodynamic sampling of the partition functions gives good results (i.e. in agreement with molecular dynamics simulations) for the temperature regime of interest in epitaxial growth experiments. The harmonic approximation is less satisfactory, but still yields the correct order of magnitude of the surface hopping rate[32]. For systems with low energy barriers ($< 3k_{\mathrm{B}}T$), or for collective diffusion processes, it is generally difficult to judge the validity of TST. In the latter case, even locating all saddle points that can be reached from a given initial state is a challenge. For this task, algorithms that allow for locating saddle points without prior knowledge of the final state have been developed. The 'dimer' method[33] is an example for such a method. It is well suited for being used together with DFT calculations, as it requires only first (spatial) derivatives of the PES, and is robust against numerical errors in the forces.

4.2 Accelerated molecular dynamics

In this Section, I'll briefly introduce methods that are suitable if the lattice approximation cannot be made, or if one needs to go beyond transition state theory. These methods employ some refinement of molecular dynamics that allows one to speed up the simulations, such that so-called 'rare' events can be observed during the run-time of a simulation. In this context, 'rare' event means an event whose rate is much smaller than the frequencies of vibrational modes. Keeping the TST estimate of rate constants in mind, any process that requires to overcome a barrier of several $k_{\mathrm{B}}T$ is a 'rare' event. Still it could take place millions of times on an experimental time scale, say, within one second. Therefore 'rare' events could be very relevant for example for simulations of epitaxial growth. Ref.[46] provides a more detailed overview of this field.

Obviously, running simulations in parallel is one possible way to access longer time scales. In the **parallel replica method**[34], one initiates several parallel simulations of the canonical ensemble starting in the same initial minimum of the PES, and observes if the system makes a transition to any other minimum. Each replica runs independently and evolves differently due to different fluctuating forces. From the abundances of various transitions observed during the parallel run, one can estimate the rate constants of these transitions, and give upper bounds for the rates of possible other transitions that did not occur during the finite runtime. The method is computationally very heavy, but has the advantage of being unbiased towards any (possibly wrong) expectations how the relevant processes may look like.

Another suggestion to speed up MD simulations goes under the term **hyperdynamics**[35]. The 'rare event problem' is overcome by adding an artificial potential to the PES that

retains the barrier region(s) but modifies the minima so as to make them shallower. The presence of such a 'boost potential' will allow the particle to escape from the minimum more quickly, and hence the processes of interest (transitions to other minima) can be observed within a shorter MD run. The method can be justified rigorously for simulations where one is interested in thermodynamic equilibrium properties (e.g., partition function): The effect of the boost potential can be corrected for by introducing a time-dependent weighting factor in the sampling of time averages. It has been suggested to extend this approach beyond thermal equilibrium to kinetical simulations: While the trajectory passes the barrier region unaffected by the boost potential, the simulation time corresponds directly to physical time. While the particle stays near a minimum of the PES, and thus under the influence of the boost potential, its effect must be corrected by making the physical time to advance faster than the simulation time. Ways to construct the boost potential in such a way that the method yields unchanged thermal averages of observables have been devised[35]. However, it has been argued that the speed-up of a simulation of epitaxy achievable with such a global boost potential is only modest if the system, as usually the case, consists of many particles[36]. This restriction can be overcome by using a local boost potential[37,38] rather than a global one. In this case it is assumed that the transitions to be 'boosted' are essentially single-particle hops. This, of course, curtails one strength of accelerated MD methods, namely being unbiased towards the (possibly wrong) expectations of the users what processes should be the important ones. Also, it is important to note that the procedure for undoing the effect of the boost potential relies on assumptions of the same type as TST. Therefore hyperdynamics cannot be used to calculate rate constants more accurately than TST.

To be able to observe more transitions and thus obtain better statistics within the (precious) runtime of an MD simulation, people have come up with the simple idea of increasing the simulation temperature. This approach is particularly attractive if one wants to simulate a physical situation where the temperature is low, e.g. low-temperature epitaxial growth of metals. By running at an artificially raised temperature (For solids, the melting temperature is an upper bound), a speed-up by several orders of magnitude may be achieved. Of course, the physics at high and low temperatures is different, thus invalidating a direct interpretation of the results obtained in this way. However, combining the idea of increased temperature MD with the principles used in kMC simulations provides us with a powerful tool. It comes under the name of **temperature-accelerated MD**[39], abbreviated as TAD: First, a bunch of MD simulations is performed, starting from the same initial state (as in the parallel replica method), at a temperature T_{high} higher than the physical temperature. The transitions observed during these runs are used for estimating their individual rates and for building up a process list. At this stage, TST in the harmonic approximation is used to 'downscale' the rate constants from their high-temperature value obtained from the MD simulation to their actual value at the lower physical temperature T_{low}. If a process is associated with an energy barrier ΔE_i, its rate constant should be scaled with a factor $\exp(\Delta E_i(T_{\mathrm{high}}^{-1} - T_{\mathrm{low}}^{-1})/k_{\mathrm{B}})$. Having sampled many MD trajectories, it is also possible to provide an upper bound for the probability that some possibly relevant transition has not yet occurred in the available set of trajectories. In other words, in TAD simulations some kind of 'safe-guarding' can be applied not to overlook possibly important transitions. After sufficiently many trajectories have been run, a pre-defined confidence level is reached that the transitions observed so far are representative for the physical behaviour of the system in

the given initial state, and can be used as a process list. Next, a kMC step is performed by selecting randomly one of the processes with probability proportional to the (scaled) rates in the process list. Then the selected process is executed, and the system's configuration changes to a new minimum. The loop is closed by going back to the first step and performing MD simulations for the system starting from this new minimum, and attempting new transitions from there.

Some more comments may be helpful. First, we note that the scaling factor used for downscaling the rates is different for different processes. Thus, the method accounts for the fact that the relative importance of high-barrier processes and low-barrier processes must be different at high and low temperatures, respectively. This requirement of a physically meaningful kinetical simulation would be violated by just naively running MD at a higher temperature without applying any corrections, but TAD passes this important test. Secondly, TAD may even provide us with information that goes beyond TST. For instance, if collective diffusion processes play a role, the relative abundance with which they were encountered in the MD runs gives us a direct estimate of the associated attempt frequency, without having to invoke the (sometimes questionable) approximations of TST. [e] Third, one can gain in computational efficiency by using the same ideas as in kMC: The process list need not be build from scratch each time, but only those entries that changed since the last step need to be updated.

Using this strategy, TAD has been applied to simulations of epitaxy[40]. In this context, it should be noted that the need for starting MD simulations in each simulation step can be reduced further: As mentioned above, kMC is based on the idea that the local environment of a particle, and thus the processes accessible for this particle, can be classified. Once the TAD simulations have established the rates for all processes of a certain environmental class (e.g. terrace diffusion), these rates can be re-used for all particles in this class (e.g., all single adatoms on a terrace). This reduces the computational workload considerably.

Finally, we mention that TAD has recently been combined with parallel kMC simulations using the semi-rigorous synchronous sublattice algorithm[41].

5 Tackling with Complexity

In the early literature of the field, kMC simulations are typically considered as a tool to rationalize experimental findings. In this approach, one works with models that are as simple as possible, i.e., comprise as few process types as possible, while still allowing for reproducing the experimental data. The rates of these processes are then often treated as parameters whose values are adjusted to fit the data. The aim is to find a description of the experimental observations with a minimum number of parameters.

More recently, the focus has shifted to kMC simulations being perceived as a scale-bridging simulation technique that enables researchers to describe a specific material or materials treatment as accurately as desired. The goal is to perform kMC simulations where *all relevant* microscopic processes are considered, aiming at simulations with predictive power. This could mean that all distinct processes derived from a given Hamiltonian, e.g. an SOS model, should be included. However, for predictive simulations, a model

[e] The assumption that TST can be used for downscaling is a milder one than assuming the applicability of TST for the attempt frequency as such.

Hamiltonian is often an already too narrow basis. The ultimate benchmark are (possibly accelerated) MD simulations that allow for an unbiased assessment which processes are relevant for a specific material, and then to match the kMC simulations to these findings.

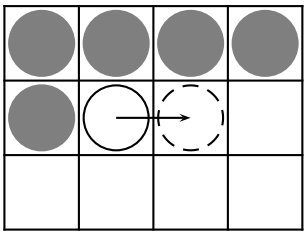


Figure 6. Illustration of the local environments of a hopping particle (white circle, in its initial (full line) and final (dashed line) state) in a model with nearest-neighbor interactions. The classification may depend on the occupation of the sites adjacent to the initial or the final state. A particular occupation is indicated by the grey circles. Since each of the ten relevant sites may be either occupied or empty, there can be up to 2^{10} environment classes.

As explained in the preceding Section, the efficiency of kMC simulations rests on a classification scheme for the local environments a particle encounters during the course of a simulation: Wherever and whenever the particle is in the 'same' environment, the same process types and rate constants will be re-used over and over again. However, the number of different process types to be considered may be rather big. For example, even for the simple SOS model, the complete process list could have up to 2^{10} entries[42] (as explained below). This raises the issue of complexity: Apart from approximations for calculating rate constants (such as TST), a kMC simulation may be more or less realistic depending on whether the classification of local environments and processes is very fine-grained, or whether a more coarse classification scheme is used.

On one end of the complexity scale, we find kMC simulations that do not rely on a pre-defined process list. Instead, the accessible processes are re-determined after each step, i.e., the process list is generated 'on the fly' while the simulation proceeds. This can be done for instance by temperature-accelerated molecular dynamics (see preceding Section). If one is willing to accept TST as a valid approximation for the calculation of rate constants, molecular dynamics is not necessarily required; instead, it is computationally more efficient to perform a **saddle point search**, using a modified dynamics for climbing 'up-hill' from a local minimum. An example for a saddle point search algorithm that uses such 'up-hill climbing' is the 'dimer' method[36]. The overall design of the kMC algorithm employing the 'search on the fly' is similar to TAD: Starting from an initial state, one initiates a bunch of saddle point searches. For each saddle point encountered, TST is used to calculate the associated rate constant. If repeated searches find the known saddle points again and again with a similar relative frequency, one can be confident that the transitions found so far make up the complete process list for this particular initial state. Next, a kMC step is carried out, leading to a new configuration; the saddle point search is continued from there, etc.

List-free kMC has been used to study metal epitaxy. For aluminum, for instance, these studies have revealed the importance of collective motion of groups of atoms for the surface

mass transport[43]. We note that the lattice approximation is not essential for this approach. Thus, it could even be applied to investigate kinetics in amorphous systems. While the saddle point search is considerably faster than MD, the method is, however, still orders of magnitude more expensive than list-directed kMC, in particular if used in conjunction with DFT to obtain the potential energy surface and the forces that enter the saddle point search.

The above method becomes computationally more affordable if the lattice approximation is imposed. The constraint that particles must sit on lattice sites reduces the possibility of collective motions, and thus invoking the lattice approximation makes the method less general. On the other hand, using a lattice makes it easier to re-use rate constants calculated previously for processes taking place in the 'same' local environment. A variant of kMC called 'self-learning'[44,45] also belongs into this context: Here, one starts with a pre-defined process list, but the algorithm is equipped with the ability to add new processes to this list if it encounters during the simulation a local environment for which no processes have been defined so far. In this case, additional saddle point searches have to be performed in order to obtain the rate constants to be added to the list.

At a lower level of complexity, we find kMC simulations that employ, in addition to the lattice approximation, a finite-range model for the interactions between particles. For the local minima of the PES, this implies that the depths of the minima can be described by a lattice Hamiltonian. For each minimum, there is an on-site energy term. If adjacent sites are occupied, the energy will be modified by pair interactions, triple interactions, etc. In materials science, this way of representing an observable in terms of the local environments of the atoms is called cluster expansion method (see the contribution by S. Müller in this book).

The usage of a lattice Hamiltonian or cluster expansion is in principle an attractive tool for tackling with the complexity in a kMC simulation of crystalline materials. However, for calculating rate constants, we need (in TST) the energy *differences* between the transition state and the initial minimum the particle is sitting in. This complicates the situation considerably. To discuss the issue, let's assume that the interactions between particles are limited to nearest neighbors. Then, both the initial state and the final state of the particle can be characterized completely by specifying which of their neighbors are occupied. On a 2D square lattice, a particle moving from one site to a (free) neighboring site has a shell of ten adjacent sites that could be either occupied or free (see Fig. 6). Thus, the move is completely specified (within the nearest-neighbor model) by one out of 2^{10} possible local environments[42]. [f] One way to specify the input for a kMC simulation is to specify a rate constant for each of these 2^{10} process types. This is in principle possible if an automated algorithm is used to determine the energy barrier and attempt frequency for each case. For practical purposes, one may specify only a selected subset of the 2^{10} rate constants, and assume that the rest takes on one of these specified values. This is equivalent to assuming that, at least for some environments, the occupation of some of the ten sites doesn't matter. This approach has been used by the author to describe the rate constants for epitaxy on a semiconductor surface, GaAs(001)[20]. A process list with about 30 entries was employed to describe the most relevant process types.

Another way of tackling with the complexity is the assumption that ΔE does not depend on the occupation of sites, but only on the *energies* of the initial and final minima. The

[f] To be precise, the actual number is somewhat smaller due to symmetry considerations.

technical advantage of this approach lies in the fact that the energies of the minima may be represented via a lattice Hamiltonian (or, equivalently, by the cluster expansion method). Thus, these energies can be retrieved easily from the cluster expansion. However, there is no rigorous foundation for such an assumption, and its application could introduce uncontrolled approximations. For a pair of initial and final states, $i = (\mathrm{ini}, \mathrm{fin})$, one could, for example, assume that $\Delta E = \Delta E_0 + \frac{1}{2}(E_{\mathrm{fin}} - E_{\mathrm{ini}})$. This assumption has been employed for diffusion of Ag adatoms on the Ag(111) surface in the presence of interactions[22], and test calculations using DFT for selected configurations of adatoms have confirmed its validity. Note that the dependence on the sum of the initial and final state energy assures that the forward and backward rate fulfill detailed balance, Eq. (17), as required for a physically meaningful simulation.

In a large part of the literature on kMC, an even simpler assumption is made, and the rate constants are assumed to depend on the energy of the initial state only. In other word, the transition states for *all* processes are assumed to be at the same absolute energy. This assumption facilitates the simulations, but clearly is not very realistic. At this point, we have reached the opposite end on the scale of complexity, where the goal is no longer a realistic modeling of materials, but a compact description of experimental trends.

I would like to conclude this Section with a word of caution: In epitaxial growth, fine details of the molecular processes may have drastic consequences on the results of the simulations. Often, the details that make a difference are beyond the description by a lattice Hamiltonian. One example is the mass transport between adjacent terraces by particles hopping across a surface step. In many metals, the energy barrier for this process is somewhat higher than the barrier for conventional hopping diffusion on the terraces. This so-called Schwöbel-Ehrlich effect is crucial for the smoothness or roughness of epitaxially grown films, but is not accounted for by the SOS model. Thus, the rate for hopping across steps needs to be added 'by hand' to the process list of the SOS model to obtain sensible simulation results. Another example concerns the shape of epitaxial islands on close-packed metal surfaces, for instance Al(111) and Pt(111). Here, either triangular or hexagonal islands can be observed, depending on the temperature at which an experiment of epitaxial growth is carried out. A detailed analysis shows that the occurrence of triangular islands is governed by the process of corner diffusion: An atom sitting at the corner of a hexagonal island, having an island atom as its only neighbor, has different probabilities for hopping to either of the two island edges[10,11]. For this reason, there is a tendency to fill up one particular edge of a hexagonal island, and the island gradually evolves to a triangular shape. Only at higher temperatures, the difference between the two rates becomes less pronounced, and the hexagonal equilibrium shape of the islands evolves. Only with the help of DFT calculations it has been possible to detect the difference of the energy barriers for the two processes of corner diffusion. Simplified models based on neighbor counting, however, cannot detect such subtle differences, in particular if only the initial state is taken into account. Therefore, kinetic Monte Carlo studies addressing morphological evolution should always be preceded by careful investigations of the relevant microscopic processes using high-level methods such as DFT for calculating the potential energy profiles.

6 Summary

With this tutorial I intended to familiarise the readers with the various tools to carry out scale-bridging simulations. These tools range from accelerated molecular dynamics simulations that extend the idea of Car-Parrinello molecular dynamics to longer time scales, to abstract models such as the lattice-gas Hamiltonian. The scientist interested in applying one of these tools should decide whether she/he wants to trust her/his intuition and start from an educated guess of a suitable kinetic model, such as SOS or similar. Else, she/he may prefer to 'play it safe', i.e. avoid as much as possible the risk of overlooking rare, but possibly important events. In the latter case, kMC simulations in combination with saddle point searches (that build up the rate list 'on the fly') are a good choice. However, this methods could be computationally too expensive if slow changes in a system very close to equilibrium should be studied, or if vastly different processes play a role whose rates span several orders of magnitude. In this case, considerations of numerical efficiency may demand from the user to make a pre-selection of processes that will be important for the evolution of the system towards the non-equilibrium structures one is interested in. Using the N-fold way kinetic Monte Carlo algorithm with a pre-defined list of process types could be a viable solution for these requirements. In summary, Monte Carlo methods allow one to go in either direction, to be as accurate as desired (by including sufficiently many many details in the simulation), or to find a description of nature that is as simple as possible.

Acknowledgments

I'd like to acknowledge helpful discussions with Wolfgang Paul, with whom I had the pleasure of organizing a joint workshop on kinetic Monte Carlo methods. Matthias Timmer is thanked for reading the manuscript and making suggestions for improvements.

References

1. R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55:2471, 1985.
2. W. Paul and J. Baschnagel. *Stochastic Processes: From Physics to Finance*. Springer, 1999.
3. D. Landau and K. Binder. *A Guide to Monte Carlo Simulation in Statistical Physics*. Cambridge University Press, 2000.
4. P. Kratzer and M. Scheffler. Surface knowledge: Toward a predictive theory of materials. *Comp. Sci. Engineering*, 3:16–25, 2001.
5. A.L. Barabasi and H. E. Stanley. *Fractal Concepts in Surface Growth*. Cambridge University Press, 1995.
6. G. H. Gilmer. Growth on imperfect crystal faces : I. Monte-Carlo growth rates. *J. Cryst. Growth*, 36:15–28, 1976.
7. S. Clarke and D. D. Vvedensky. Origin of reflection high-energy electron-diffraction intensity oscillations during molecular-beam epitaxy: A computational modeling approach. *Phys. Rev. Lett.*, 58:2235, 1987.

8. T. Michely and J. Krug. *Islands, Mounds and Atoms – Patterns and Processes in Crystal Growth Far from Equilibrium*, volume 42 of *Springer Series in Surface Science*. Springer, 2004.
9. Z. Zhang and M. G. Lagally. Atomistic processes in the early stages of thin-film growth. *Science*, 276:377, 1997.
10. A. Bogicevic, J. Strömquist, and B. I. Lundqvist. Low-symmetry diffusion barriers in homoepitaxial growth of Al(111). *Phys. Rev. Lett.*, 81:637–640, 1998.
11. S. Ovesson, A. Bogicevic, and B. I. Lundqvist. Origin of compact triangular islands in metal-on-metal growth. *Phys. Rev. Lett.*, 83:2608–2611, 1999.
12. M. Kotelyanskii and D. N. Theodorou, editors. *Simulation Methods for Polymers*. CRC Publishers, 2004.
13. M. Neurock and E. W. Hansen. First-principles-based molecular simulation of heterogeneous catalytic surface chemistry. *Computers in Chemical Engineering*, 22(Suppl.):S1045 – S1060, 2000.
14. L. J. Broadbelt and R. Q. Snurr. Applications of molecular modeling in heterogeneous catalysis research. *Applied Catalysis A*, 200:23 – 46, 2000.
15. K. Reuter and M. Scheffler. First-principles kinetic Monte Carlo simulations for heterogeneous catalysis: Application to the CO oxidation at RuO_2(110). *Phys. Rev. B*, 73:045433, 2006.
16. C. Sendner, S. Sakong, and A. Groß. Kinetic Monte Carlo simulations of the partial oxidation of methanol on oxygen-covered Cu(110). *Surf. Sci.*, 600:3258 – 3265, 2006.
17. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes*. Cambridge University Press, 1986.
18. A. B. Bortz, M. H. Kalos, and J. L. Lebowitz. New algorithm for Monte Carlo simulations of Ising spin systems. *J. Comp. Phys.*, 17:10–18, 1975.
19. K. A. Fichthorn and W. H. Weinberg. Theoretical foundations of dynamical Monte Carlo simulations. *J. Chem. Phys.*, 95:1090, 1991.
20. P. Kratzer and M. Scheffler. Reaction-limited island nucleation in molecular beam epitaxy of compound semiconductors. *Phys. Rev. Lett.*, 88:036102, 2002.
21. P. Kratzer, E. Penev, and M. Scheffler. First-principles studies of kinetics in epitaxial growth of III-V semiconductors. *Appl. Phys. A*, 75:79–88, 2002.
22. K. A. Fichthorn and M. Scheffler. Island nucleation in thin-film epitaxy: A first-principles investigation. *Phys. Rev. Lett.*, 84:5371, 2000.
23. P. A. Maksym. Fast Monte Carlo simulation of MBE growth. *Semicond. Sci. Technol.*, 3:594, 1988.
24. J. J. Lukkien, J. P. L. Segers, P. A. J. Hilbers, R. J. Gelten, and A. P. J. Jansen. Efficient Monte Carlo methods for the simulation of catalytic surface reactions. *Phys. Rev. E*, 58:2598, 1998.
25. B. Lehner, M. Hohage, and P. Zeppenfeld. Novel Monte Carlo scheme for the simulation of adsorption and desorption processes. *Chem. Phys. Lett.*, 336:123, 2001.
26. Y. Shim and J. G. Amar. Rigorous synchronous relaxation algorithm for parallel kinetic Monte Carlo simulations of thin film growth. *Phys. Rev. B*, 71:115436, 2005.
27. M. Merrik and K. A. Fichthorn. Synchronous relaxation algorithm for parallel kinetic Monte Carlo simulations of thin film growth. *Phys. Rev. E*, 75:011606, 2007.
28. Y. Shim and J. G. Amar. Semirigorous synchronous sublattice algorithm for parallel

kinetic Monte Carlo simulations of thin film growth. *Phys. Rev. B*, 71:125432, 2005.
29. P. Hänggi, P. Talkner, and M. Borkovec. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.*, 62(2):251–341, April 1990.
30. S. Glasstone, K. J. Laidler, and H. Eyring. *The Theory of Rate Processes*. McGraw-Hill, New York, 1941.
31. G. H. Vineyard. Frequency factors and isotope effects in solid state rate processes. *J. Phys. Chem. Solids*, 3:121, 1957.
32. G. Boisvert, N. Mousseau, and L. J. Lewis. Surface diffusion coefficients by thermodynamic integration: Cu on Cu(100). *Phys. Rev. B*, 58:12667, 1998.
33. G. Henkelman and H. Jónsson. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivates. *J. Chem. Phys.*, 111:7010, 1999.
34. A. F. Voter. Parallel replica method for dynamics of infrequent events. *Phys. Rev. B*, 78:3908, 1998.
35. A. F. Voter. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.*, 78:3908, 1997.
36. G. Henkelman and H. Jónsson. Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. *J. Chem. Phys.*, 115:9657–9666, 2001.
37. J.-C. Wang, S. Pal, and K. A. Fichthorn. Accelerated molecular dynamics of rare events using the local boost method. *Phys. Rev. B*, 63:085403, 2001.
38. R. A. Miron and K. A. Fichthorn. Heteroepitaxial growth of Co/Cu(001) : An accelerated molecular dynamics simulation study. *Phys. Rev. B*, 72:035415, 2005.
39. M. R. Sørensen and Arthur F. Voter. Temperature accelerated dynamics for simulation of infrequent events. *J. Chem. Phys.*, 112:9599, 2000.
40. F. Montalenti, M. R. Sørensen, and Arthur F. Voter. Closing the gap between experiment and theory: Crystal growth by temperature accelerated dynamics. *Phys. Rev. Lett.*, 87:126101, 2001.
41. Y. Shim, J. G. Amar, B. P. Uberuaga, and A. F. Voter. Reaching extended length scales and time scales in atomistic simulations via spatially parallel temperature-accelerated dynamics. *Phys. Rev. B*, 76:205439, 2007.
42. Arthur F. Voter. Classically exact overlayer dynamics: Diffusion of rhodium clusters on Rh(100). *Phys. Rev. B*, 34:6819, 1986.
43. G. Henkelmann and H. Jónsson. Multiple time scale simulations of metal crystal growth reveal the importance of multiatom surface processes. *Phys. Rev. Lett.*, 90:116101, 2003.
44. O. Trushin, A. Karim, A. Kara, and T. S. Rahman. Self-learning kinetic Monte Carlo method: Application to Cu(111). *Phys. Rev. B*, 72:115401, 2005.
45. A. Karim, A. N. Al-Rawi, A. Kara, T. S. Rahman, O. Trushin, and T. Ala-Nissila. Diffusion of small two-dimensional Cu islands on Cu(111) studied with a kinetic Monte Carlo method. *Phys. Rev. B*, 73:165411, 2006.
46. A. F. Voter, F. Montalenti, and T. C. Germann. Extending the time scale in atomistic simulation of materials. *Annu. Rev. Mater. Res.*, 32:321–346, 2002.

Electronic Structure: Hartree-Fock and Correlation Methods

Christof Hättig

Lehrstuhl für Theoretische Chemie
Fakultät für Chemie und Biochemie
Ruhr-Universität Bochum, 44780 Bochum, Germany
E-mail: christof.haettig@rub.de

Hartree-Fock theory is the conceptually most basic electronic structure method and also the starting point for almost all wavefunction based correlation methods. Technically, the Hartree-Fock self-consistent field method is often also the starting point for the development of molecular Kohn-Sham density functional theory codes. We will briefly review the main concepts of Hartree-Fock theory and modern implementations of the Rothaan-Hall self-consistent field equations with emphasis on the techniques used to make these approaches applicable to large systems. The second part of the chapter will focus on wavefunction based correlation methods for large molecules, in particular second order Møller-Plesset perturbation theory (MP2) and, for calculations on excited states, the approximate coupled-cluster singles-and-doubles method CC2, both treating the electron-electron interaction correct through second order. It is shown how the computational costs (CPU time and storage requirements) can be reduced for these methods by orders of magnitudes using the resolution-of-the-identity approximation for electron repulsion integrals. The demands for the auxiliary basis sets are discussed and it shown how these approaches can parallelized for distributed memory architectures. Finally a few prototypical applications are reviewed.

1 Introduction

Today, essentially all efficient electronic structure methods are based on the Born-Oppenheimer approximation and molecular orbital theory. The Hartree-Fock method combines these two concepts with the variation principle and the simplest possible wave function ansatz obeying the Pauli exclusion principle: a Slater determinant or, for open-shell systems in restricted Hartree-Fock theory, a configuration state function. In spite of the fact that Hartree-Fock is since decades a matured quantum-chemical method, its implementation for large scale application is still today an active field of research. The reason for this is not that there is a large interest in the results from the Hartree-Fock calculations themselves. The driving force behind these developments are today the technical similarity between Hartree-Fock (HF) theory and Kohn-Sham density functional theory (DFT), in particular if hybrid functionals are used, and the fact that Hartree-Fock calculations are the starting point for almost all wavefunction based correlation methods. The challenge for HF and DFT implementations is today an efficient prescreening of the numerical important contributions and the storage of sparse matrices in large scale parallel calculations.

During the last decade also many wavefunction based correlation methods have been proposed for applications on extended molecular systems. Most of them are based on the so-called local correlation approach[1–9], and/or on an extensive screening of small but often long ranging contributions to the correlation energy[4,10,11]. Some approaches introduce empirical parameters or rely on a balance between certain contributions which in practice might or might not be given[12–15]. For most of these approaches it is not yet clear to

which extend they can be developed in the near future into competitive methods for extended systems. In particular, if the reduction of the computational costs (compared to more traditional implementations of quantum chemical methods) relies on a screening in the atomic orbital (AO) basis set, calculations on extended systems are often only possible with rather small basis sets which cannot supply the accuracy expected from a correlated *ab initio* method.[a] Even though usually only explored for electronic ground states, most of these approaches could in principle also be generalized to excited states. But for larger molecules, calculations for excited states employ often so-called response methods and the parameterization of the excited state used in these methods hampers the application of the local correlation and related approaches.[16–18]

We will in the following not go into the details of these approaches, but restrict ourself to discussion of to the underlying electronic structure methods, which are usually single-reference coupled-cluster (CC) and, in particular, for larger systems Møller-Plesset perturbation theory through second order (MP2) or related methods for excited states. The implementation of the latter methods has during the last decade improved dramatically by combining them with the so-called resolution-of-the-identity (RI) approximation for the four-index electron repulsion integrals (ERIs) with optimized auxiliary basis sets. Even without any further approximations are these methods today applicable to systems with up to 100 or more atoms. Since the RI approximation depends little on the electronic structure of the investigated system it does not diminish the applicability of the underlying electronic structure methods. It is also compatible and can be combined with the above mentioned screening based approaches to reduce further the computational costs.[19,20] Thus, it can be expected that these two aspects, the treatment of the electron correlation through second order and the RI approximation for ERIs will remain important ingredients also in future correlated wavefunction based methods for extended systems.

In the following the theory of wavefunction based ab initio methods that treat the electron-electron interaction correctly through second order is briefly reviewed. The emphasis will be on methods for excited states which can be related to the approximate coupled-cluster singles-and-doubles model CC2, an approximation to the coupled-cluster singles-and-doubles method (CCSD). In Sec. 7 it is shown how the computational costs for these methods can be reduced drastically by using the RI approximation and disc space bottlenecks for these methods can be resolved by an doubles amplitudes-direct implementation. A recent parallel implementation for distributed memory architectures is presented in Sec. 8 and some example applications with RI-MP2 and RI-CC2 are reviewed in Secs. 9 and 10.

2 The Born-Oppenheimer Approximation and the Electronic Schrödinger Equation

An important simplification in the quantum mechanical description of molecules, which is ubiquitously applied in electronic structure calculations is the Born-Oppenheimer (BO) approximation which leads to a separation of the electronic from the nuclear degrees of

[a] Here and in the following we use "ab initio" for electronic structure methods which are systematically improvable in the sense that they are members of a hierarchy which converges to the exact solution of the electronic Schrödinger equations, i.e. the full configuration interaction (Full CI) limit.

freedom. In the BO approximation the total Hamiltonian of molecular system is split in the operator for the kinetic energy $\hat{T}_{nuc}$ of the nuclei and the remaining contributions which are put into an electronic Hamiltonian $\hat{H}_{el}$.

$$\hat{H}_{tot} = \hat{T}_{nuc} + \hat{H}_{el} \tag{1}$$

In the non-relativistic case we have

$$\hat{T}_{nuc} = -\sum_{A} \frac{1}{2M_A} \hat{\nabla}_A^2 \tag{2}$$

and the electronic Hamiltonian can be written in atomic units as

$$\hat{H}_{el}(\mathbf{r}, \mathbf{R}) = -\sum_{i} \frac{1}{2} \hat{\nabla}_i^2 - \sum_{i,A} \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{A<B} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \,, \tag{3}$$

where $\hat{\nabla}_A$ and $\hat{\nabla}_i$ are the gradients with respect to the coordinates of nucleus A and electron i, respectively, $\mathbf{R}_A$ and $\mathbf{r}_i$, and Z_A the charge of nucleus A.

The total wavefunction is approximated as product of an electronic and a nuclear wavefunction

$$\Psi_{tot}(\mathbf{r}, \mathbf{R}) \approx \Psi_{el}(\mathbf{r}, \mathbf{R}) \Psi_{nuc}(\mathbf{R}) \,. \tag{4}$$

where the electronic wavefunction is determined as eigenfunction of the electronic Hamiltonian

$$\hat{H}_{el}(\mathbf{r}, \mathbf{R}) \Psi_{el}(\mathbf{r}, \mathbf{R}) = E_{el}(\mathbf{R}) \Psi_{el}(\mathbf{r}, \mathbf{R}) \,, \tag{5}$$

and the nuclear wavefunction as solution of a nuclear Schrödinger equation

$$\Big(\hat{T}_{nuc} + E_{el}(\mathbf{R})\Big) \Psi_{nuc}(\mathbf{r}, \mathbf{R}) = E_{tot} \Psi_{nuc}(\mathbf{r}, \mathbf{R}) \,, \tag{6}$$

in which the eigenvalues of the electronic Hamiltonian, $E_{el}(\mathbf{R})$, appear as potential for the nuclear motion. It is therefore that we speak of $E_{el}(\mathbf{R})$ as potential energy surfaces. Our understanding of molecular structures as equilibrium positions on potential energy surfaces are implicit results of the Born-Oppenheimer approximation.

One may ask, what are the errors of the BO approximation? Beside the simplified wavefunction ansatz, Eq. (4), one neglects the so-called non-adiabatic coupling elements[b]:

$$\mathbf{A}^{(A)}(\vec{R}) = \int \Psi_{el}(\mathbf{r}, \mathbf{R})^* \Big(\hat{\nabla}_A \Psi_{el}(\mathbf{r}, \mathbf{R})\Big) d\mathbf{r} \tag{7}$$

$$B^{(A)}(\vec{R}) = \int \Psi_{el}(\mathbf{r}, \mathbf{R})^* \Big(\hat{\nabla}_A^2 \Psi_{el}(\mathbf{r}, \mathbf{R})\Big) d\mathbf{r} \tag{8}$$

There appear if the total Hamiltonian is applied to Ψ_{tot},

$$\begin{aligned} \hat{H}_{tot}\Psi_{tot} &= \hat{H}_{el}\Psi_{el}(\mathbf{r}, \mathbf{R})\Psi_{nuc}(\mathbf{R}) + \Psi_{el}(\mathbf{r}, \mathbf{R})\hat{T}_{nuc}\Psi_{nuc}(\mathbf{R}) \\ &\quad - \sum_{A} \frac{1}{2M_A} \Big\{ 2\Big(\hat{\nabla}_A \Psi_{el}(\mathbf{r}, \mathbf{R}\Big) \cdot \Big(\hat{\nabla}_A \Psi_{nuc}(\mathbf{R})\Big) + \Big(\hat{\nabla}_A^2 \Psi_{el}(\mathbf{r}, \mathbf{R}\Big) \Psi_{nuc}(\mathbf{R}) \Big\} \\ &= E_{tot}\Psi_{tot} \,, \end{aligned} \tag{9}$$

[b] Note that $\mathbf{A}^{(A)}(\vec{R})$ is a three-dimensional vector in the coordinate space of nucleus A, while $B^{(A)}(\vec{R})$ is scalar.

after integration over the electronic degrees of freedom:

$$\hat{H}_{tot}\Psi_{nuc} = \left(\hat{T}_{nuc} + E_{el}(\vec{R})\right)\Psi_{nuc} \qquad (10)$$
$$-\sum_A \frac{1}{2M_A}\left\{2\mathbf{A}^{(A)}(\mathbf{R})\cdot\hat{\nabla}_A + B^{(A)}(\mathbf{R})\right\}\Psi_{nuc} = E_{tot}\Psi_{nuc}$$

The nuclear Schrödinger equation in Eq. (6) is obtained by neglecting in the last equation the non-adiabatic coupling elements $\mathbf{A}^{(A)}(\mathbf{R})$ and $B^{(A)}(\mathbf{R})$. The errors introduced by the Born-Oppenheimer approximation are typically in the order of 0.1 kJ/mol and for the majority of applications today completely negligible compared to other errors made in the solution of the electronic and nuclear Schrödinger equations, Eqs. (5) and (6).

3 Slater Determinants

The Pauli principle requires that the electronic wavefunction Ψ_{el} is antisymmetric under any permutation of two electrons i and j,

$$\hat{P}_{ij}\Psi_{el}(\mathbf{r}_1,\ldots,\mathbf{r}_i,\ldots,\mathbf{r}_j,\ldots) = \Psi_{el}(\mathbf{r}_1,\ldots,\mathbf{r}_j,\ldots,\mathbf{r}_i,\ldots) \qquad (11)$$
$$= -\Psi_{el}(\mathbf{r}_1,\ldots,\mathbf{r}_i,\ldots,\mathbf{r}_j,\ldots)\,.$$

The simplest ansatz fulfilling this condition are Slater determinants, antisymmetrized products of one-electron wavefunctions (orbitals):

$$\Psi_{SD} = \tfrac{1}{\sqrt{n}}\hat{\mathcal{A}}\psi_1(\mathbf{r}_1)\ldots\psi_n(\mathbf{r}_n) = \begin{vmatrix} \psi_1(\mathbf{r}_1) & \psi_1(\mathbf{r}_2) & \ldots & \psi_1(\mathbf{r}_n) \\ \psi_2(\mathbf{r}_1) & \psi_2(\mathbf{r}_2) & \ldots & \psi_2(\mathbf{r}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_n(\mathbf{r}_1) & \psi_n(\mathbf{r}_2) & \ldots & \psi_n(\mathbf{r}_n) \end{vmatrix} \qquad (12)$$

The non-symmetrized orbital products are also known as Hartree products and will in the following be denoted by Θ.

$$\Psi_{SD} = \tfrac{1}{\sqrt{n}}\hat{\mathcal{A}}\Theta \qquad \text{with} \quad \Theta(\mathbf{r}_1,\ldots,\mathbf{r}_n) = \psi_1(\mathbf{r}_1)\ldots\psi_n(\mathbf{r}_n) \qquad (13)$$

The antisymmetrizer $\hat{\mathcal{A}}$ is defined as

$$\hat{\mathcal{A}} = \sum_{m=1}^{n!}\mathrm{sign}(P_m)\hat{P}_m \qquad (14)$$

where $\hat{P}_m$ is an operator which performs one of the $n!$ possible permutations of the n electrons and $\mathrm{sign}(P_m)$ the parity of this permutation. The group permutation operators has the property that if the whole set of all $n!$ possible permutations of n elements $\{\hat{P}_1,\hat{P}_2,\ldots,\hat{P}_{n!}\}$ is multiplied with some permutation $\hat{P}_k$ the same set of operators is recovered, just in a different order[c]:

$$\{\hat{P}_k\hat{P}_1,\hat{P}_k\hat{P}_2,\ldots,\hat{P}_k\hat{P}_{n!}\} = \{\hat{P}_1,\hat{P}_2,\ldots,\hat{P}_{n!}\}\,. \qquad (15)$$

Furthermore, the permutation operators $\hat{P}_m$ are unitary

$$\hat{P}_m^\dagger = \hat{P}_m^{-1} \qquad (16)$$

[c]This relation is in group theory known as rearrangement theorem.

where $\hat{P}_m^{-1}$ is the operator which performs the inverse permutation which has the same parity as P_m, i.e. $\mathrm{sign}(P_m) = \mathrm{sign}(P_m^{-1})$ and

$$\{\hat{P}_1^{-1}, \hat{P}_2^{-1}, \ldots, \hat{P}_{n!}^{-1}\} = \{\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_{n!}\} \,. \tag{17}$$

From these relations it follows that the antisymmetrizer $\hat{\mathcal{A}}$ is an hermitian operator

$$\hat{\mathcal{A}}^\dagger = \sum_m \mathrm{sign}(P_m)\hat{P}_m^\dagger = \sum_m \mathrm{sign}(P_m^{-1})\hat{P}_m^{-1} = \hat{\mathcal{A}} \,, \tag{18}$$

and

$$\hat{\mathcal{A}}^2 = n!\hat{\mathcal{A}} \,. \tag{19}$$

Both relations are useful for evaluating matrix elements (integrals) for slater determinants. In the following we skip for convenience the index el for the electronic Hamiltonian and write it as

$$\hat{H} = E_{nuc} + \sum_i \hat{h}_i + \sum_{i<j} \frac{1}{r_{ij}} \tag{20}$$

with the nuclear repulsion energy and the one-electron hamiltonian defined as

$$E_{nuc} = \sum_{A<B} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \,, \tag{21}$$

and

$$\hat{h}_i = -\frac{1}{2}\hat{\nabla}_i^2 - \sum_A \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}_i|} \,, \tag{22}$$

and the interelectronic distances $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. Note that, because the summations are over all electrons or electron pairs, the antisymmetrizer $\hat{\mathcal{A}}$ commutes separately with the one- and two-electron contributions to the Hamiltonian $\hat{H}$:

$$\sum_i \hat{h}_i \hat{\mathcal{A}} = \hat{\mathcal{A}} \sum_i \hat{h}_i \qquad \text{and} \qquad \sum_{i<j} \frac{1}{r_{ij}} \hat{\mathcal{A}} = \hat{\mathcal{A}} \sum_{i<j} \frac{1}{r_{ij}} \tag{23}$$

For operators of this form we can rewrite the matrix elements for Slater determinants as

$$\begin{aligned} \big\langle \Psi_{SD,I} \big| \hat{O} \big| \Psi_{SD,J} \big\rangle &= \big\langle \tfrac{1}{\sqrt{n!}} \hat{A}\Theta_I \big| \hat{O} \big| \tfrac{1}{\sqrt{n!}} \hat{A}\Theta_J \big\rangle = \tfrac{1}{n!} \big\langle \hat{A}\Theta_I \big| \hat{A}\hat{O} \big| \Theta_J \big\rangle \\ &= \tfrac{1}{n!} \big\langle \hat{A}^2\Theta_I \big| \hat{O} \big| \Theta_J \big\rangle = \big\langle \hat{A}\Theta_I \big| \hat{O} \big| \Theta_J \big\rangle \end{aligned} \tag{24}$$

The results have, however, only a simple form if the orbitals ψ_i are orthogonal to each other. We will therefore in the following without loss of generality assume that Θ_I and Θ_J are build from a common set of orthonormal orbitals

$$\langle \psi_i | \psi_j \rangle = \delta_{ij} \tag{25}$$

and that the orbitals are ordered in the Hartree products according to increasing indices:

$$\Theta_I = \psi_{I_1}(\mathbf{r}_1)\psi_{I_2}(\mathbf{r}_2)\ldots\psi_{I_n}(\mathbf{r}_n) \qquad \text{with } I_1 < I_2 < \ldots < I_n \tag{26}$$

Overlap integrals then become

$$\left\langle \hat{A}\Theta_I \middle| \Theta_J \right\rangle = \sum_{m=1}^{n!} \mathrm{sign}(P_m) \prod_{k=1}^{n} \langle \psi_{I_{P_m(k)}} | \psi_{J_k} \rangle = \sum_{m=1}^{n!} \mathrm{sign}(P_m) \prod_{k=1}^{n} \delta_{I_{P_m(k)}, J_k} \tag{27}$$

where $P_m(k)$ is the result at position k after applying the permutation P_m because of Eq. (26) only the identity permutation can contribute to the result which is nonzero only if in both Hartree products exactly the same orbitals are occupied. We thus find that

$$\left\langle \Psi_{SD,I} \middle| \Psi_{SD,J} \right\rangle = \left\langle \hat{A}\Theta_I \middle| \Theta_J \right\rangle = \delta_{I,J} \tag{28}$$

Similarly, one obtains for the matrix elements of one-electron operators:

$$\left\langle \hat{A}\Theta_I \middle| \sum_i \hat{h}_i \middle| \Theta_J \right\rangle = \sum_{m=1}^{n!} \mathrm{sign}(P_m) \sum_{i=1}^{n} \langle \psi_{I_{P_m(i)}} | \hat{h}_i | \psi_{J_i} \rangle \prod_{\substack{k=1 \\ k \neq i}}^{n} \langle \psi_{I_{P_m(k)}} | \psi_{J_k} \rangle \, . \tag{29}$$

For an orthonormal orbital basis the matrix elements between two Slater determinants thus become:

$$\left\langle \Psi_{SD,I} \middle| \sum_i \hat{h}_i \middle| \Psi_{SD,J} \right\rangle = \begin{cases} \sum_{k=1}^{n} \langle \psi_{I_k} | \hat{h}_k | \psi_{I_k} \rangle & \text{for } I = J \\ \langle \psi_k | \hat{h} | \psi_l \rangle & \text{if } \Psi_{SD,I}, \Psi_{SD,J} \text{ differ only in } \psi_k, \psi_l \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

Nonvanishing matrix elements are obtained if the two Slater determinants are identical or differ at most in one orbital. The matrix elements for the two-electron operators become:

$$\begin{aligned} \left\langle \Psi_{SD,I} \middle| \sum_{i<j} \tfrac{1}{r_{ij}} \middle| \Psi_{SD,J} \right\rangle &= \left\langle \hat{A}\Theta_I \middle| \sum_{i<j} \tfrac{1}{r_{ij}} \middle| \Theta_J \right\rangle \\ &= \sum_{m=1}^{n!} \mathrm{sign}(P_m) \sum_{i<j}^{n} \langle \psi_{I_{P_m(i)}} \psi_{I_{P_m(j)}} | \tfrac{1}{r_{ij}} | \psi_{J_i} \psi_{J_j} \rangle \\ &\quad \times \prod_{\substack{k=1 \\ k \neq i,j}}^{n} \langle \psi_{I_{P_m(k)}} | \psi_{J_k} \rangle \, , \end{aligned} \tag{31}$$

which reduces for orthonormal orbitals to:

$$\left\langle \Psi_{SD,I} \middle| \sum_{i<j} \tfrac{1}{r_{ij}} \middle| \Psi_{SD,J} \right\rangle = \begin{cases} \sum_{k<l}^{n} \langle \psi_{I_k} \psi_{I_l} || \psi_{I_k} \psi_{I_l} \rangle & \text{for } I = J \\ \sum_{m} \langle \psi_k \psi_{I_m} || \psi_l \psi_{I_m} \rangle & \text{if } \Psi_{SD,I}, \Psi_{SD,J} \text{ differ only in } \psi_k, \psi_l \\ \langle \psi_i \psi_j || \psi_k \psi_l \rangle & \text{if } \Psi_{SD,I}, \Psi_{SD,J} \text{ differ in } \psi_i, \psi_k \text{ and } \psi_j, \psi_m \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

For two-electron operators non-vanishing matrix elements are obtained for Slater determinants which differ in up to two orbitals. The antisymmetrized integrals introduced on right side of Eq. (32) are defined as:

$$\langle \psi_i \psi_j || \psi_k \psi_l \rangle = (\psi_i \psi_k | \psi_j \psi_l) - (\psi_i \psi_l | \psi_j \psi_k) \,, \tag{33}$$

with the two-electron integrals in the Mulliken notation given by

$$(\psi_i \psi_j | \psi_k \psi_l) = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \psi_i^*(\mathbf{r}_1) \psi_j(\mathbf{r}_1) \frac{1}{r_{12}} \psi_k^*(\mathbf{r}_2) \psi_l^*(\mathbf{r}_2) \, d\mathbf{r}_1 \, d\mathbf{r}_2 \,. \tag{34}$$

The expectation value of the total electronic Hamiltonian for a Slater determinant with the orthonormal occupied orbitals $\psi_1, \ldots, \psi_n$ is thus given by:

$$\big\langle \Psi_{SD} \big| \hat{H} \big| \Psi_{SD} \big\rangle = E_{nuc} + \sum_i \big\langle \psi_i \big| \hat{h} \big| \psi_i \big\rangle + \frac{1}{2} \sum_{ij} \big\langle \psi_i \psi_j \big|\big| \psi_i \psi_j \big\rangle \,. \tag{35}$$

4 Hartree-Fock Theory and the Roothaan-Hall Equations

The basic idea behind Hartree-Fock theory is to take the simplest meaningful ansatz for the electronic wavefunction, a Slater determinant, and to determine the occupied orbitals by the variation principle, i.e. such that energy expectation value is minimized. For general molecular or extended systems this scheme is usually combined with a basis set expansion of the molecular orbitals.

$$\psi_i(1) = \sum_{\nu=1}^{N} \chi_\nu(\mathbf{r}_1) C_{\nu i} \cdot \begin{cases} \alpha(1) \\ \beta(1) \end{cases} \,, \tag{36}$$

where $\{\chi_\nu\}$ is a basis set with N spatial functions and α und β are spin function for, respectively, the "spin up" and "spin down" states. For extended systems often plane wave basis sets is used, but for molecular systems local atom centered basis sets (*linear combination of atomic orbitals*, LCAO) are more common.

To minimize the Hartree-Fock energy with respect to the MO coefficients $c_{\nu i}$ under the constraint that the ψ_i are orthonormal we introduce the Lagrange function,

$$L_{\mathrm{HF}} = E_{nuc} + \sum_i \big\langle i \big| \hat{h} \big| i \big\rangle + \frac{1}{2} \sum_{ij} \big\langle ij \big|\big| ij \big\rangle + \sum_{ij} \epsilon_{ji} \big(\delta_{ij} - \big\langle i \big| j \big\rangle \big) \,. \tag{37}$$

Here and in the following we skip for notational convenience the functions ψ and χ in the brackets and give only there indices with the convention that $i, j, \ldots$ denote occupied MOs and greek indices AOs. The Lagrange function L_{HF} is now required to be stationary with respect to arbitrary variations of the MO coefficients[d]:

$$\frac{dL_{\mathrm{HF}}}{dC_{\nu i}^*} = \big\langle \nu \big| \hat{h} \big| i \big\rangle + \sum_k \big\langle \nu k \big|\big| i k \big\rangle - \sum_j \epsilon_{ji} \big\langle \nu \big| j \big\rangle = 0 \,. \tag{38}$$

[d] Requiring the derivatives $dL_{\mathrm{HF}}/dC_{\nu i}$ to vanish leads to equivalent complex conjugated equations.

We now introduce the Fock and overlap matrices in atomic orbital basis $\{\chi_\nu\}$ as:

$$F_{\mu\nu} = \big\langle \mu \big| \hat{h} \big| \nu \big\rangle + \sum_k \big\langle \mu k \big|\big| \nu k \big\rangle \tag{39}$$

$$= \big\langle \mu \big| \hat{h} \big| \nu \big\rangle + \sum_{\kappa\lambda} D_{\kappa\lambda} \Big\{ \big(\mu\nu\big|\kappa\lambda\big) - \big(\mu\lambda\big|\kappa\nu\big) \Big\} , \tag{40}$$

and

$$S_{\mu\nu} = \big\langle \mu \big| \nu \big\rangle , \tag{41}$$

with AO density matrix $\mathbf{D}$ defined as:

$$D_{\kappa\lambda} = \sum_k C^*_{\kappa k} C_{\lambda k} . \tag{42}$$

Note that the Hartree-Fock energy can be calculated from the Fock and densities matrices and the matrix of elements of the one-electron hamiltonian $h_{\mu\nu} = \langle \mu | \hat{h} | \nu \rangle$ as

$$E_{\mathrm{HF}} = \frac{1}{2} \sum_{\mu\nu} D_{\mu\nu} \Big(F_{\mu\nu} + h_{\mu\nu} \Big) . \tag{43}$$

With these intermediates Eq. (38) can be rewritten in a compact matrix form:

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\epsilon} . \tag{44}$$

The last equation is known under the name "Roothaan-Hall equation". Its meaning becomes more clear if it is transformed to an orthonormal basis set

$$\tilde{\chi}_\mu = \sum_\nu \chi_\nu [\mathbf{S}^{-1/2}]_{\nu\mu} \qquad \text{with} \quad \mathbf{S}^{-1/2}\mathbf{S}^{-1/2} = \mathbf{S}^{-1} , \tag{45}$$

where $[\mathbf{S}^{-1/2}]_{\mu\nu}$ denotes[e] the element μ, ν of the matrix $\mathbf{S}^{-1/2}$. In this basis the Roothaan-Hall equations become

$$\sum_\nu \tilde{F}_{\mu\nu} \tilde{C}_{\nu i} = \sum_j \tilde{C}_{\mu j} \epsilon_{ji} \qquad \text{with} \quad \tilde{F} = \mathbf{S}^{-1/2}\mathbf{F}\mathbf{S}^{-1/2} \quad \text{and} \quad \tilde{C} = \mathbf{S}^{1/2}\mathbf{C} . \tag{46}$$

The result of the Fock matrix applied any occupied orbital is a linear combination of only occupied orbitals. This condition determines the occupied molecular orbitals only up to a unitary transformation of these orbitals among themselves, which leaves the Slater determinant, i.e. the Hartree-Fock wavefunction, unchanged.

The so-called canonical orbitals are obtained by choosing this unitary transformation such that the matrix with the lagrangian multipliers ϵ_{ji} becomes diagonal. Usually, the equation is then augmented by a similar condition for the complementary space of unoccupied or "virtual" orbitals. The Roothaan-Hall equations become then a generalized nonlinear eigenvalue problem—nonlinear since the Fock matrix $\mathbf{F}$ depends through the density matrix $\mathbf{D}$ on the solution of the equations. The standard algorithm to solve these equations is the self-consistent field procedure which can be sketched as follows:

1. Initially a start density matrix is guessed (or constructed from some start orbitals, e.g. from an extended Hückel calculation)

[e] Note that $[\mathbf{S}^{-1/2}]_{\mu\nu} \neq 1/\sqrt{S_{\mu\nu}}$.

2. The Fock matrix F and the total energy for the approximate density matrix are calculated using Eqs. (39) and (43).

3. The generalized eigenvalue problem Eq. (44) is solved to obtain a new set of MOs.

4. An improved density matrix is guessed from the present approximation for the MOs and the previous density matrices using some convergence acceleration procedure.

5. If the total energy, the MOs and the density are converged (i.e. self-consistent) the procedure is stopped, else one continuous with step 2.

The number of iterations needed to converge the self-consistent field procedure depends on the molecular system (in particular its HOMO-LUMO gap), the quality of the start guess and a lot on the method used to update the density matrix in step 4. A common choice is the direct inversion of iterative subspaced (DIIS) technique of Pulay[21,22].

5 Direct SCF, Integral Screening and Integral Approximations

Apart from the technique used to solve the Roothaan-Hall equations, i.e. to update the density matrix, a second technically demanding aspect is the construction of the Fock matrix. A naive implementation of Eq. (39) would require the calculation of $\approx \frac{1}{8}N^4$ two-electron integral, where N is the dimension of our basis set in Eq. (36). To achieve a useful accuracy, typically 10–30 basis functions are needed per atom. For many systems of interest in computational chemistry today with 100 and more atoms the number of two-electron integrals will even today exceed standard disc space capacities. Furthermore, a brute force summation over all integrals would be unnecessary costly in terms of CPU time: for local atom-center basis sets many of the two-electron integrals and, depending on the HOMO-LUMO gap, also of the density matrix are numerically negligible; in extended systems the number of numerically significant two-electron coulomb integrals will only grow with $\mathcal{O}(\mathcal{N}^2)$, where $\mathcal{N}$ is a measure of the system size. A solution to these problems is offered by the integral-direct SCF scheme in combination with integral prescreening:

- The two-electron integrals are not stored once on stored on file, but instead (re)calculated when needed and immediately contracted with the elements of the density matrix to increments of the Fock matrix. By exploiting the eightfold permutational symmetry

$$(\mu\nu|\kappa\lambda) = (\nu\mu|\kappa\lambda) = (\mu\nu|\lambda\kappa) = (\nu\mu|\lambda\kappa) = \tag{47}$$
$$(\kappa\lambda|\mu\nu) = (\kappa\lambda|\nu\mu) = (\lambda\kappa|\mu\nu) = (\lambda\kappa|\nu\mu) \tag{48}$$

  of the two-electron integrals, one can restrict the loop over the AO indices to $\mu < \nu$, and $\kappa < \lambda$ with $(\mu, \nu) < (\kappa, \lambda)$ and add for each two-electron integral the following

6 increments to the Fock matrix:

$$F_{\mu\nu} \leftarrow F_{\mu\nu} + 2D_{\kappa\lambda}(\mu\nu|\kappa\lambda) \tag{49}$$
$$F_{\kappa\lambda} \leftarrow F_{\kappa\lambda} + 2D_{\mu\nu}(\mu\nu|\kappa\lambda) \tag{50}$$
$$F_{\mu\lambda} \leftarrow F_{\mu\lambda} - D_{\nu\kappa}(\mu\nu|\kappa\lambda) \tag{51}$$
$$F_{\nu\lambda} \leftarrow F_{\nu\lambda} - D_{\mu\kappa}(\mu\nu|\kappa\lambda) \tag{52}$$
$$F_{\mu\kappa} \leftarrow F_{\mu\kappa} - D_{\nu\lambda}(\mu\nu|\kappa\lambda) \tag{53}$$
$$F_{\nu\kappa} \leftarrow F_{\nu\kappa} - D_{\mu\lambda}(\mu\nu|\kappa\lambda) \tag{54}$$

(Where we assumed for simplicity that all four AO indices are different, else the redundant increments have to be skipped.)

- To estimate whether a specific integral might be large enough to make a significant contribution to the Fock matrix one exploits e.g. the Schwarz condition[f]:

$$|(\mu\nu|\kappa\lambda)| \leq Q_{\mu\nu}Q_{\kappa\lambda} \qquad \text{with} \quad Q_{\mu\nu} = \sqrt{(\mu\nu|\mu\nu)}\,. \tag{55}$$

For a given index quadruple the integral $(\mu\nu|\kappa\lambda)$ needs only to be calculated if

$$Q_{\mu\nu}Q_{\kappa\lambda}D_{max} \geq \tau \tag{56}$$

where

$$D_{max} = \max\{2|D_{\mu\nu}|, 2|D_{\kappa\lambda}|, |D_{\nu\kappa}|, |D_{\mu\kappa}|, |D_{\nu\lambda}|, |D_{\mu\lambda}|\}\,, \tag{57}$$

and τ is a user-defined threshold that determines the numerical accuracy of the calculation. Only if the inequality is fulfilled any of the contributions to the Fock matrix in Eqs. (49) – (54) can become larger than the threshold τ. This technique is today standard in essentially all direct Hartree-Fock codes and also in molecular DFT codes for so-called Hybrid functional with an Hartree-Fock-like "exact exchange" contribution.

For large systems the integral-screening reduces the computational costs for the Fock matrix construction from $\mathcal{O}(\mathcal{N}^4)$ to $\mathcal{O}(\mathcal{N}^2)$. If we split the two-electron part of the Fock matrix into separate Coulomb and exchange contributions,

$$F_{\mu\nu} = h_{\mu\nu} + J_{\mu\nu} - K_{\mu\nu}\,, \tag{58}$$

with

$$J_{\mu\nu} = \sum_{\kappa\lambda} D_{\kappa\lambda}(\mu\nu|\kappa\lambda)\,, \qquad \text{and} \qquad K_{\mu\nu} = \sum_{\kappa\lambda} D_{\kappa\lambda}(\mu\lambda|\kappa\nu)\,, \tag{59}$$

the remaining $\mathcal{O}(\mathcal{N}^2)$ scaling is caused by the Coulomb contribution while for the exchange part the integral screening reduces the number of requires contributions asymptotically to $\mathcal{O}(\mathcal{N})$ if the HOMO-LUMO gap does not vanish and the density matrix becomes sparse. This becomes more clear if the parameter D_{max} for the Coulomb and exchange contributions to the Fock matrix are calculated separately:

$$\text{Coulomb:} \qquad D_{max,C} = \max\{2|D_{\mu\nu}|, 2|D_{\kappa\lambda}|\} \tag{60}$$
$$\text{exchange:} \qquad D_{max,X} = \max\{|D_{\nu\kappa}|, |D_{\mu\kappa}|, |D_{\nu\lambda}|, |D_{\mu\lambda}|\} \tag{61}$$

[f] The Schwarz condition for two-electron integrals is a special case of a Cauchy-Schwarz inequality for scalar products in vector space: $|\langle x, y\rangle| \leq ||x|| \cdot ||y||$ with $||x|| = \sqrt{\langle x, x\rangle}$.

The size of the absolute values of the density matrix elements $D_{\mu\nu}$ and of the quantities $Q_{\mu\nu}$ are correlated with the overlap of the basis functions χ_μ and χ_ν. Thus, $D_{max,C}$ becomes usually only small if also the integral $|(\mu\nu|\kappa\lambda)| \leq Q_{\mu\nu} \cdot Q_{\kappa\lambda}$ is small, while in the exchange case the density matrix elements contributing to $D_{max,X}$ have indices then the Q's and criterion $Q_{\mu\nu}Q_{\kappa\lambda}D_{max,X}$ will only be fulfilled if all four basis functions μ, ν, κ, and λ are close in space.

Also, for medium sized molecules or with basis sets which contain diffuse functions only modest computational savings obtained with this technique and the large costs for the individual two-electron integrals can hamper the applicability of Hartree-Fock self-consistent field calculations. An approximation which leads to a significant reduction of the computational costs for the Coulomb contribution to the Fock matrix construction is the resolution-of-the-identity approximation for the two-electron integrals which is also known as density fitting:

$$(\mu\nu|\kappa\lambda) \approx (\mu\nu|Q)\,[\mathbf{V}^{-1}]_{QP}\,(P|\kappa\lambda)\;, \tag{62}$$

where $(\mu\nu|Q)$ and V_{PQ} are, respectively, three- and two-center two-electron integrals:

$$(\mu\nu|Q) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \chi^*_\mu(\mathbf{r}_1)\chi_\nu(\mathbf{r}_2)\frac{1}{r_{12}}Q(\mathbf{r}_2)\,d\mathbf{r}_1 d\mathbf{r}_2\;, \tag{63}$$

$$V_{PQ} = (Q|P) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} P(\mathbf{r}_1)\frac{1}{r_{12}}Q(\mathbf{r}_2)\,d\mathbf{r}_1 d\mathbf{r}_2\;. \tag{64}$$

Within this approximation the Coulomb matrix $J_{\mu\nu}$ can be calculated as:

$$\gamma_P = \sum_{\kappa\lambda}(P|\kappa\lambda)D_{\kappa\lambda} \tag{65}$$

$$\sum_Q V_{PQ}c_Q = \gamma_P \tag{66}$$

$$J_{\mu\nu} \approx \sum_Q (\mu\nu|Q)c_Q \tag{67}$$

Where Eq. (66) is linear equation system for c_Q. In combination with an integral screening based on the Schwarz inequality these three equations can also be implemented with an asymptotic scaling of $\mathcal{O}(\mathcal{N}^2)$, but a significant lower prefactor than the original method, since there are fewer two- and three-center two-electron integrals and the computational costs for them are lower than for the four-center two-electron integrals $(\mu\nu|\kappa\lambda)$. We optimized auxiliary basis sets $\{Q\}$, which are today available for several standard basis sets, the errors introduced by the RI approximation are insignificant compared to the basis incompleteness error of the LCAO expansion in Eq. (36).

6 Second Order Methods for Ground and Excited States

Second order Møller-Plesset perturbation theory is a conceptually simple and technically the most simplest ab initio correlation method. It can be derived by expanding the solution of the electronic Schrödinger equation as a Taylor series in the fluctuation potential (vide infra). This can be done either in the framework of configuration interaction theory or using the single-reference coupled-cluster ansatz for the wavefunction.[23] We will take here

the latter starting point to have a close connection to coupled-cluster response and related methods for excited states. In the coupled-cluster ansatz the wavefunction is parameterized as

$$|\mathrm{CC}\rangle = \exp(\hat{T})|\mathrm{HF}\rangle \tag{68}$$

with the cluster operator defined as

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots \tag{69}$$

where

$$\hat{T}_1 = \sum_{\mu_1} t_{\mu_1} \hat{\tau}_{\mu_1} = \sum_{ai} t_a^i \hat{\tau}_a^i \;, \quad \hat{T}_2 = \sum_{\mu_2} t_{\mu_2} \hat{\tau}_{\mu_2} = \sum_{aibj} t_{ab}^{ij} \hat{\tau}_{ab}^{ij} \;, \quad \dots \tag{70}$$

The coefficients t_{μ_i} are called cluster amplitudes and the excitation operators $\hat{\tau}_{\mu_i}$ generate all possible single, double, and higher excited determinants if applied on the ground state Hartree-Fock (HF) determinant $|\mathrm{HF}\rangle$. Here and in the following, we use the convention that indices i, j, ... denote occupied, a, b, ... virtual, and p, q, ... arbitrary molecular orbitals (MOs).

Inserting the ansatz (68) into the electronic Schrödinger equation and multiplying from the left with $\exp(-\hat{T})$ one gets

$$\exp(-\hat{T})\hat{H}\exp(\hat{T})|\mathrm{HF}\rangle = E|\mathrm{HF}\rangle \;. \tag{71}$$

Projecting the above form of the Schrödinger equation onto the HF determinant and a projection manifold of (suitable linear combinations of) excited determinants one obtains an expression for the ground state energy

$$E = \langle \mathrm{HF}|\exp(-\hat{T})\hat{H}\exp(\hat{T})|\mathrm{HF}\rangle = \langle \mathrm{HF}|\hat{H}\exp(\hat{T})|\mathrm{HF}\rangle \;, \tag{72}$$

and the cluster equations

$$0 = \langle \mu_i|\exp(-\hat{T})\hat{H}\exp(\hat{T})|\mathrm{HF}\rangle \;, \tag{73}$$

which determine the amplitudes t_{μ_i}. Since we have not yet made any approximation, the above equations still give the exact ground state solution of the electronic Schrödinger equation. Truncating the cluster operator (69) after the single ($\hat{T}_1$) and double ($\hat{T}_2$) excitations gives the coupled-cluster singles-and-doubles (CCSD) method, truncating it after $\hat{T}_3$ the CCSDT method, and so on.[g]

Expressions for Møller-Plesset perturbation theory are found by splitting the Hamiltonian into the Fock operator $\hat{F}$ as zeroth-order and the electron-electron fluctuation potential as first-order contribution to the Hamiltonian

$$\hat{H}^{(0)} = \hat{F} \;, \qquad \hat{H}^{(1)} = \hat{\Phi} = \hat{H} - \hat{F} \;, \tag{74}$$

and expanding Eqs. (72) and (73) in orders of the fluctuation potential. If the Brillouin-Theorem is fulfilled and $\langle_a^i|\hat{H}|\mathrm{HF}\rangle = 0$, i.e. for a closed-shell or an unrestricted open-shell Hartree-Fock (UHF) reference, the MP2 energy is obtained as

$$E_{\mathrm{MP2}} = \langle \mathrm{HF}|\hat{\Phi}\hat{T}_2^{(1)}|\mathrm{HF}\rangle = \sum_{aibj} t_{ab}^{ij} \langle \mathrm{HF}|\hat{\Phi}|_{ij}^{ab}\rangle \tag{75}$$

[g] Similar as in configuration interaction theory, a truncation after single excitations (CCS) does not give a useful method for the calculation of ground state energies. As follows from the Brillouin theorem $\langle_a^i|\hat{H}|\mathrm{HF}\rangle = 0$, the cluster equations have then for a closed-shell or an unrestricted open-shell reference determinant the trivial solution $t_a^i = 0$ and the CCS energy becomes equal the HF energy.

with

$$0 = \langle^{ab}_{ij}|[\hat{F}, \hat{T}_2^{(1)}] + \hat{\Phi}|\mathrm{HF}\rangle \quad \Leftrightarrow \quad t^{ij}_{ab} = \frac{\langle^{ab}_{ij}|\hat{\Phi}|\mathrm{HF}\rangle}{\epsilon_i - \epsilon_a + \epsilon_j - \epsilon_b} \tag{76}$$

where we assumed canonical molecular orbitals and ϵ_p are the orbital energies.

Møller-Plesset perturbation theory can not straightforwardly be applied to excited states, since wavefunctions for excited states usually require a multi-reference treatment. For reviews on multi-reference many-body perturbation theory and its application on electronically excited states see e.g. Refs. 24, 25. Correlated second order methods for the calculation of excitation energies based on a single-reference treatment for electronic ground states can, however, be derived within the framework of coupled-cluster response theory. The idea behind response theory is to study a system exposed to time-dependent external (e.g. electric) fields and to derive from the response of the wavefunction or density the frequency-dependent properties of the system—for example polarizabilities and hyperpolarizabilities. The latter properties have singularities whenever a frequency of a field becomes equal to the excitation energy of an allowed transition in the system. Thus, from the poles of frequency-dependent properties one can identify the excitation energies.

Consider a quantum mechanical system described in the unperturbed limit by the time-independent Hamiltonian[h] $\hat{H}^{(0)}$ which is perturbed by a time-dependent potential:

$$\hat{H}(t, \epsilon) = \hat{H}^{(0)} + \hat{V}(t, \epsilon) \, . \tag{77}$$

We assume that the perturbation $\hat{V}$ can be expanded as a sum over monochromatic Fourier components

$$\hat{V}(t, \epsilon) = \sum_j \hat{V}_j \epsilon_j e^{-i\omega_j t} \, , \tag{78}$$

where $\hat{V}_j$ are hermitian, time-independent one-electron operators (e.g. for an electric field the dipole operator), t the time and ϵ_j are the amplitudes of the associated field strengths. Then the full time-dependent wavefunction of the system, i.e. the solution to the time-dependent Schrödinger equation, can be expanded as a power series in the field strengths as

$$\Psi(t) = \underbrace{\left[\Psi^{(0)} + \sum_j \Psi_j^{(1)}(\omega_j)\epsilon_j e^{-i\omega_j t} + \dots\right]}_{\text{phase-isolated wavefunction } \tilde{\Psi}} e^{-i\int_{t_0}^{t} dt' \langle\tilde{\Psi}(t)|\hat{H}(t',\epsilon) - i\frac{\partial}{\partial t'}|\tilde{\Psi}(t)\rangle} \, , \tag{79}$$

and an expectation value for an operator $\hat{\mu}$ as

$$\langle\mu\rangle(t) = \langle\tilde{\Psi}(t)|\hat{\mu}|\tilde{\Psi}(t)\rangle = \mu^{(0)} + \sum_j \langle\langle\mu; V_j\rangle\rangle_{\omega_j} \epsilon_j e^{-i\omega_j t} + \dots \tag{80}$$

For detailed reviews of modern response theory and its implementation for approximate wavefunction methods the interested reader is referred to Refs. 26–31. The important point for the calculation excitation energies is that the poles in the response functions $\langle\langle\mu; V\rangle\rangle_\omega$ occur when ω becomes equal to an eigenvalue of the stability matrix of the employed

[h] Note that $\hat{H}^{(0)}$ includes here the fluctuation potential in difference to Eq. (74), where the fluctuation potential $\hat{\Phi}$ has been the perturbation.

electronic structure method for the unperturbed system. The stability matrix contains the derivatives of the residua of the equations which determine the wavefunction parameters with respect to these parameters. For Hartree-Fock, multi-configurational self-consistent field (MCSCF), density functional theory (DFT), configuration interaction (CI) or other methods which are variational in the sense that the wavefunction parameters are determined by minimization of the energy, the stability matrix is the so-called electronic Hesse matrix—the matrix of the second derivatives of the energy with respect to the wavefunction parameters. For coupled-cluster methods the cluster amplitudes are determined by the cluster equations (73). Arranging the residua in a vector function

$$\Omega_{\mu_i}(t_{\nu_i}) = \langle \mu_i | \exp(-\hat{T}) \hat{H} \exp(\hat{T}) | \mathrm{HF} \rangle \ , \tag{81}$$

the stability matrix is given by the Jacoby matrix

$$\mathbf{A}_{\mu_i \nu_j} = \left. \frac{d\Omega_{\mu_i}}{dt_{\nu_j}} \right|_{\epsilon=0} = \langle \mu_i | \exp(-\hat{T}) [\hat{H}, \hat{\tau}_{\nu_j}] \exp(\hat{T}) | \mathrm{HF} \rangle \ , \tag{82}$$

where $|_{\epsilon=0}$ indicates that the derivatives are taken for the unperturbed system, i.e. at zero field strengths. In configuration interaction theory the stability matrix becomes the matrix representation of the reduced Hamiltonian $\hat{H} - E_0$ (where E_0 is the ground state energy) in the space orthogonal to the electronic ground state.[i] In coupled-cluster theory this matrix representation is obtained in a similarity transformed basis.[j]

In this way excitation energies can in principle be derived for any electronic structure method. However, to obtain physical meaningful and accurate results, the method has to fulfill certain requirements. For example from the equations for the amplitudes in MP2, Eq. (76), one obtains a Jacoby matrix which gives only excitation energies corresponding to double excitations and these would be equal to the orbital energy differences in the denominator of the amplitudes. The two most important requirements are firstly, that there must be a one-to-one correspondence between the parameters of the wavefunction and at least the investigated part of the spectrum of the Hamiltonian. This requires methods which determine the time-dependent variables by a single set of equations, as e.g. time-dependent Hartree-Fock (HF-SCF), density functional theory (DFT) or multi-configuration self-consistent field (MCSCF, CASSCF, or RASSCF), but not a time-dependent configuration interaction (CI) treatment on top of a time-dependent HF-SCF calculation. For this reason the coefficients of the Hartree-Fock orbitals are also above in Eqs. (81) and (82) not considered as parameters of the time-dependent wavefunction, since this second set of variables in the time-dependent problem would lead to a second set of eigenvalues corresponding to single excited states, additionally to the one obtained from the parameterization through the singles cluster amplitudes. Instead, the time-dependent wavefunction is in coupled-cluster response theory usually constructed using the (time-independent) orbitals of the unperturbed system with time-dependent cluster amplitudes. Secondly, to obtain accurate results the stability matrix must also provide an accurate approximation of the those blocks of the Hamiltonian which are most important for the investigated states. For single excitations these are the singles-singles block $\mathbf{A}_{\mu_1 \nu_1}$ and the off-diagonal blocks

[i] In connection with CI and propagator methods (approximate) matrix representations of $\hat{H} - E_0$ are often also referred to as secular matrix.

[j] $\langle \mu_i | \exp(-\hat{T})$ for the bra and $\exp(\hat{T}) | \mu_j \rangle$ for the ket states, where $|\mu_j\rangle = \hat{\tau}_{\mu_j} |\mathrm{HF}\rangle$; for further details see e.g. Ref. 23

$\mathbf{A}_{\mu_2\nu_1}$ and $\mathbf{A}_{\mu_1\nu 2}$ next to it. With the usual single-reference coupled-cluster methods these blocks are described most accurately and therefore the excitation energies for single excitation dominated transitions are obtained with the highest accuracy, while excitation energies for double and higher excitations are usually considerably less accurate.

Already at the coupled-cluster singles (CCS) level (which for excitation energies is—in contrast to ground state calculations—not equivalent to Hartree-Fock, but to configuration interaction singles (CIS)), excitation energies for states dominated by single replacements of one spin-orbital in the Hartree-Fock reference determinant are obtained correctly through first order in the electron-electron interaction.

A second order method for excited states which accounts for the above requirements and takes over the accuracy of MP2 to excited states dominated by single excitations can be derived by approximating the cluster equations to lowest order in the fluctuation potential. But in difference to the derivation of MP2 in Eqs. (74) – (76) we allow in the Hamiltonian for an additional one-electron perturbation

$$\hat{H}(t) = \hat{F} + \hat{\Phi} + \hat{V}(t) \,, \tag{83}$$

which can induce transitions to single excitations and has, as necessary in CC response theory, not been included in the Hartree-Fock calculation. Because of the latter, single excitation amplitudes contribute now to the cluster operator already in zeroth order in the fluctuation potential, $\hat{\Phi}$, and in first order $\hat{T}_1$ and $\hat{T}_2$ both contribute to the wavefunction. Approximating the equations that determine these amplitudes to second (singles) and first order (doubles) one obtains the equations for the approximate coupled-cluster model CC2[32,33]:

$$0 = \langle {}^a_i|[\hat{\tilde{H}}, \hat{T}_2] + \hat{\tilde{H}}|\mathrm{HF}\rangle \,, \tag{84}$$

$$0 = \langle {}^{ab}_{ij}|[\hat{F}, \hat{T}_2] + \hat{\tilde{H}}|\mathrm{HF}\rangle \,, \tag{85}$$

where a similarity transformed Hamiltonian $\hat{\tilde{H}} = \exp(-\hat{T}_1)\hat{H}\exp(\hat{T}_1)$ has been introduced to obtain a compact notation. In difference to MP2 the equations for CC2 have to be solved iteratively because of the coupling introduced by $\hat{T}_1$. The ground state energy obtained from CC2

$$E_{\mathrm{CC2}} = \langle \mathrm{HF}|\hat{\Phi}(\hat{T}_2 + \tfrac{1}{2}\hat{T}_1\hat{T}_1)|\mathrm{HF}\rangle \,, \tag{86}$$

is, as for MP2, (only) correct through second order in the fluctuation potential[k], but it leads to a Jacoby matrix with the singles-singles block $\mathbf{A}_{\mu_1\nu_1}$ correct through second order and the off-diagonal blocks $\mathbf{A}_{\mu_1\nu_2}$ and $\mathbf{A}_{\mu_2\nu_1}$ correct through first-order in the fluctuation potential, while the doubles-doubles $\mathbf{A}_{\mu_2\nu_2}$ block is approximated by the zeroth-order term:

$$\mathbf{A}^{\mathrm{CC2}} = \begin{pmatrix} \langle {}^a_i|[(\hat{\tilde{H}} + [\hat{\tilde{H}}, \hat{T}_2]), \hat{\tau}^c_k|\mathrm{HF}\rangle & \langle {}^a_i|[\hat{\tilde{H}}, \hat{\tau}^{cd}_{kl}]|\mathrm{HF}\rangle \\ \langle {}^{ab}_{ij}|[\hat{\tilde{H}}, \hat{\tau}^c_k]|\mathrm{HF}\rangle & \langle {}^{ab}_{ij}|[\hat{F}, \hat{\tau}^{cd}_{kl}]|\mathrm{HF}\rangle \end{pmatrix} \,. \tag{87}$$

CC2 is the computational simplest iterative coupled-cluster model which gives single excitation energies which are correct through second order. Through the similarity transformed

[k] Therefore, CC2 does in general not describe ground state energies, structures, or properties more accurately than MP2. Its advantage upon MP2 is that, combined with coupled-cluster response theory, it can (in contrast to the latter) applied successfully to excited states.

Hamiltonian $\hat{\tilde{H}} = \exp(-\hat{T}_1)\hat{H}\exp(\hat{T}_1)$ the Jacoby matrix in Eq. (87) includes, however, also some higher-order terms, since for the unperturbed system the single excitation amplitudes t_{μ_1} contribute only in second- and higher orders to the ground state wavefunction.[l] Excluding these terms and replacing the doubles amplitudes by the first-order amplitudes, Eq. (76), from which the MP2 energy in calculated, one obtains the Jacoby matrix of the CIS(D_∞) approximation[37], an iterative variant of the perturbative doubles correction[38] CIS(D) to CIS (or CCS):

$$\mathbf{A}^{\mathrm{CIS(D_\infty)}} = \begin{pmatrix} \langle {}^a_i|[(\hat{H}+[\hat{H},\hat{T}_2]),\hat{\tau}^c_k|\mathrm{HF}\rangle & \langle {}^a_i|[\hat{H},\hat{\tau}^{cd}_{kl}]|\mathrm{HF}\rangle \\ \langle {}^{ab}_{ij}|[\hat{H},\hat{\tau}^c_k]|\mathrm{HF}\rangle & \langle {}^{ab}_{ij}|[\hat{F},\hat{\tau}^{cd}_{kl}]|\mathrm{HF}\rangle \end{pmatrix} . \tag{88}$$

This Jacobian contains the minimal number of terms required to obtain the excitation energies for single replacement dominated transitions correct through second order. However, it is not possible to construct a coupled-cluster model which leads exactly to such a Jacoby matrix.

The computational savings of CIS(D_∞) compared to CC2 are rather limited[37] and CC2 has, as a member of the hierarchy of coupled-cluster methods CCS, CC2, CCSD, CC3, CCSDT, ... certain conceptual advantages. The Jacoby matrix of the CIS($D\infty$) approximation may, however, used as starting point to derive the perturbative doubles correction CIS(D) to the CIS (or CCS) excitation energies[37]:

$$\omega^{(\mathrm{D})} = \sum_{\mu_1\nu_1} E^{\mathrm{CIS}}_{\mu_1} \left[\mathbf{A}^{\mathrm{CIS(D_\infty)}}_{\mu_1\nu_1} - \mathbf{A}^{\mathrm{CIS}}_{\mu_1\nu_1} + \sum_{\kappa_2} \frac{\mathbf{A}^{\mathrm{CIS(D_\infty)}}_{\mu_1\kappa_2}\mathbf{A}^{\mathrm{CIS(D_\infty)}}_{\kappa_2\nu_1}}{\omega^{\mathrm{CIS}} - \epsilon_{\kappa_2}} \right] E^{\mathrm{CIS}}_{\nu_1} \tag{89}$$

or

$$\omega^{\mathrm{CIS(D)}} = \omega^{\mathrm{CIS}} + \omega^{(\mathrm{D})} = \sum_{\mu_1\nu_1} E^{\mathrm{CIS}}_{\mu_1} \left[\mathbf{A}^{\mathrm{CIS(D_\infty)}}_{\mu_1\nu_1} + \sum_{\kappa_2} \frac{\mathbf{A}^{\mathrm{CIS(D_\infty)}}_{\mu_1\kappa_2}\mathbf{A}^{\mathrm{CIS(D_\infty)}}_{\kappa_2\nu_1}}{\omega^{\mathrm{CIS}} - \epsilon_{\kappa_2}} \right] E^{\mathrm{CIS}}_{\nu_1} \tag{90}$$

where ϵ_{κ_2} contains the orbital energy difference for a double excitation, $\epsilon^{ij}_{ab} = \epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j$.

Another second order method for excited states which is related to CC2 and CIS(D) is the so-called algebraic diagrammatic construction through second order, ADC(2).[39,40] The secular matrix of ADC(2) is just the symmetric part of $\mathbf{A}^{\mathrm{CIS(D_\infty)}}$:

$$\mathbf{A}^{\mathrm{ADC(2)}} = \tfrac{1}{2}\mathbf{A}^{\mathrm{CIS(D_\infty)}} + \tfrac{1}{2}\left(\mathbf{A}^{\mathrm{CIS(D_\infty)}}\right)^\dagger , \tag{91}$$

which leads to some conceptual and also computational simplifications e.g. in the calculation of derivatives (gradients!) since the left and right eigenvectors of a symmetric matrix are identical, while for the non-symmetric Jacoby matrices of CC2 and CIS(D_∞) left and right eigenvectors differ. Both eigenvectors are needed for the calculation of derivatives. Other second order methods for excited states are the second order polarization propagator approach,[41,42] SOPPA and the perturbative doubles correction,[43] RPA(D), to time-dependent Hartree-Fock, which for excitation energies is also known as the random phase

[l] We assume here that the Brillouin theorem is fulfilled and thus the occupied/virtual block of the Fock matrix vanishes. This holds for closed-shell and unrestricted open-shell Hartree-Fock reference states. For a discussion of additional terms that need to be accounted for in restricted open-shell SCF based calculations we refer e.g. to Refs.[34–36].

approximation (RPA). The latter method can also be understood as a non-iterative approximation to SOPPA, similar as CIS(D) is a non-iterative approximation to CIS(D_∞). The relation of RPA(D) and SOPPA to the single-reference coupled-cluster response methods is somewhat more difficult, since these methods are members of a different hierarchy of methods (with RPA (TDHF) as first-order model) which is related to the so-called orbital-optimized coupled-cluster (OCC) methods[44,45]. Therefore, these methods will not be discussed in detail in the following, but we note that the same concepts (doubles amplitude-direct formulation and RI-approximation) can by applied to reduce also for these the computational costs to the same extend as for CC2, ADC(2), CIS(D_∞), and CIS(D).

6.1 Doubles amplitude-direct formulation of second order methods

An important feature of second order methods or approximate doubles methods, as one might also call them, is that an explicit storage (in RAM or on disk) of complete sets of double excitation amplitudes can be avoided similar as the storage of triples amplitudes is avoided in the approximate triples methods CCSD(T), CCSDT-1, CCSDR(3), or CC3.[46–49] This is important for applications on large molecules since similar as for the approximate triples methods the storage of the amplitudes would prohibit large-scale applications simply by a storage space or I/O bottleneck.

For example, the MP2 energy can be calculated without storing the double excitation amplitudes using the following scheme[m]:

```
do i = 1, nocc
  do j = i, nocc
    do a = 1, nvirt
      do b = b, nvirt
        t^ij_ab = (ia|jb)/(ε_i − ε_a + ε_j − ε_b)
        E_MP2 = E_MP2 + (2 − δ_ij){2(ia|jb) − (ia|jb)}t^ij_ab
      end do
    end do
  end do
end do
```

In a similar way also the equations for the doubles amplitudes in CC2 can—for given singles amplitudes t_a^i—immediately be inverted to

$$t_{ab}^{ij} = (ai\tilde{|}bj)/(\epsilon_i - \epsilon_a + \epsilon_j - \epsilon_b) \tag{92}$$

where the similarity transformation with $\exp(\hat{T}_1)$ has been included in the AO-to-MO transformation for the modified two-electron integrals

$$(ai\tilde{|}bj) = \sum_\alpha \Lambda^p_{\alpha a} \sum_\beta \Lambda^h_{\beta i} \sum_\gamma \Lambda^p_{\gamma b} \sum_\delta \Lambda^h_{\delta j}\ (\alpha\beta|\gamma\delta) \tag{93}$$

with $\Lambda^p_{\alpha a} = C_{\alpha a} - \sum_k C_{\alpha k} t_a^k$ and $\Lambda^h_{\alpha i} = C_{\alpha i} + \sum_c C_{\alpha c} t_c^i$. Inserting Eq. (92) into the equation for the singles amplitudes, Eq. (84), gives a set of effective equations for the CC2

[m] The explicit formulas given here and below are for a closed-shell restricted Hartree-Fock reference determinant.

singles amplitudes, which reference the doubles amplitudes t_{ab}^{ij} only as intermediates, which can be calculated and contracted with one- and two-electron integrals "on-the-fly" without storing a complete set of these amplitudes on disk:

```
do i = 1, nocc
  do j = 1, nocc
    do a = 1, nvirt
      do b = 1, nvirt
        t_ab^ij = (ai~|bj)/(ε_i - ε_a + ε_j - ε_b)
        Ω_ci = Ω_ci + Σ_abj (2t_ab^ij - t_ba^ij)(jb~|ca)
        Ω_ak = Ω_ak - Σ_bij (2t_ab^ij - t_ba^ij)(jb~|ik)
        ⋮
      end do
    end do
  end do
end do
```

To avoid the storage of doubles amplitudes is even more important for excited states, since in this case else doubles contributions to eigen- or trial vectors would have to be stored for several simultaneously solved eigenvalues and a number of iterations. An explicit reference to the doubles part of eigen- or trial vectors during the solution of the eigen problem can for the approximate doubles methods be removed by exploiting the particular structure of the Jacoby or secular matrices of these methods, in which the doubles-doubles block is in the canonical orbital basis diagonal with the diagonal elements equal to SCF orbital energy differences:

$$\begin{pmatrix} \mathbf{A}_{\mu_1\nu_1} & \mathbf{A}_{\mu_1\nu_2} \\ \mathbf{A}_{\mu_2\nu_1} & \delta_{\mu_2\nu_2}\epsilon_{\nu_2} \end{pmatrix} \begin{pmatrix} E_{\nu_1} \\ E_{\nu_2} \end{pmatrix} = \omega \begin{pmatrix} E_{\nu_1} \\ E_{\nu_2} \end{pmatrix} . \tag{94}$$

The doubles part of the eigenvectors is thus related to the singles part and the eigenvalue through the equation

$$E_{\mu_2} = \frac{\sum_{\nu_1} \mathbf{A}_{\mu_2\nu_1} E_{\nu_1}}{\omega - \epsilon_{\mu_2}} . \tag{95}$$

which allows to partition the linear eigenvalue problem in the space of singles and doubles replacements as an effective eigenvalue problem in the space of only the single excitations:

$$\sum_{\nu_1} \left[\mathbf{A}_{\mu_1\nu_1} + \sum_{\kappa_2} \frac{\mathbf{A}_{\mu_1\kappa_2}\mathbf{A}_{\kappa_2\nu_1}}{\omega - \epsilon_{\kappa_2}} \right] E_{\nu_1} = \sum_{\nu_1} \mathbf{A}^{eff}_{\mu_1\nu_1}(\omega) E_{\nu_1} = \omega E_{\mu_1} . \tag{96}$$

The last equation is, however, in difference to Eq. (94) a nonlinear eigenvalue problem because the effective Jacoby matrix $\mathbf{A}^{eff}_{\mu_1\nu_1}(\omega)$ depends on the eigenvalue ω, which is itself first known when the equation has been solved. But with iterative techniques this eigenvalue problem can be solved almost as efficiently as the original linear eigenvalue problem

and the elimination of the need to store the doubles part of solution or trial vectors more than compensates this complication.[50]

To apply these iterative techniques for the solution of large-scale eigenvalue problems one needs to implement matrix vector products of the form

$$\sigma_{\mu_1}(\omega, b_{\nu_1}) = \sum_{\nu_1} \mathbf{A}^{eff}_{\mu_1\nu_1}(\omega) b_{\nu_1} = \sum_{\nu_1} \mathbf{A}_{\mu_1\nu_1} b_{\nu_1} + \sum_{\kappa_2} \mathbf{A}_{\mu_1\kappa_2} \frac{\sum_{\nu_1} \mathbf{A}_{\kappa_2\nu_1} b_{\nu_1}}{\omega - \epsilon_{\kappa_2}} \,. \quad (97)$$

Note the similarity of the quotient in the last term with the expression in Eq. (95). For CC2 this term becomes

$$b^{ij}_{ab} = \frac{1}{\epsilon_{iajb}} \sum_{ck} \mathbf{A}_{iajb,kc} b^k_c = \frac{\sum_{ck} \langle {}^{ab}_{ij} | [\hat{\tilde{H}}, \hat{\tau}^k_c] | \mathrm{HF} \rangle b^k_c}{\epsilon_i - \epsilon_a + \epsilon_j - \epsilon_b + \omega} = \frac{2(ai|bj) - (bi|aj)}{\epsilon_i - \epsilon_a + \epsilon_j - \epsilon_b + \omega} \,, \quad (98)$$

with the modified MO electron repulsion integrals

$$(ai|bj) = \hat{P}^{ij}_{ab} \sum_{\alpha\beta} \left(\bar{\Lambda}^p_{\alpha a} \Lambda^h_{\beta i} + \Lambda^p_{\alpha a} \bar{\Lambda}^h_{\beta i} \right) \sum_{\gamma\delta} \Lambda^p_{\gamma b} \Lambda^h_{\delta j} \, (\alpha\beta|\gamma\delta) \,, \quad (99)$$

where $\bar{\Lambda}^p_{\alpha a} = -\sum_k C_{\alpha k} b^k_a$, $\bar{\Lambda}^h_{\alpha i} = +\sum_c C_{\alpha c} b^i_c$ and $\hat{P}^{ij}_{ab}$ a symmetrization operator defined through $\hat{P}^{ij}_{ab} f_{ia,jb} = f_{ia,jb} + f_{jb,ia}$. The linear transformation in Eq. (97) can thus be calculated using a similar algorithm as for the residuum of the ground state cluster equations without storing any doubles vectors:

```
do i = 1, nocc
  do j = 1, nocc
    do a = 1, nvirt
      do b = 1, nvirt
        b^ij_ab = (ai|bj)/(ε_i − ε_a + ε_j − ε_b + ω)
        σ_ci = σ_ci + Σ_abj (2 b^ij_ab − b^ij_ba)(jb|ca)
        ⋮
        t^ij_ab = (ai|bj)/(ε_i − ε_a + ε_j − ε_b)
        σ_ai = σ_ai + Σ_bj (2 t^ij_ab − t^ij_ba) Σ_ck [2(jb|kc) − (jc|kb)] b_ck
      end do
    end do
  end do
end do
```

The fact that the doubles amplitudes of CC2 are determined by the singles amplitudes through Eqs. (92) and (93) and reduce for $t_{\mu_1} \rightarrow t^{(1)}_{\mu_1} = 0$ to the first-order amplitudes of MP2, opens a simple possibility to implement CIS(D_∞) and CIS(D) as approximations to CC2. Considering the effective Jacoby matrix, Eq. (96), as a functional of the singles amplitudes $\mathbf{A}^{eff}(t_{\mu_1}, \omega)$ one obtains the connection:

$$\text{CC2} \quad : \quad \sum_{\nu_1} \mathbf{A}^{eff}_{\mu_1\nu_1}(t^{\mathrm{CC2}}_{\kappa_1}, \omega) E_{\nu_1} = \omega E_{\mu_1}$$

$$\text{CIS(D}_\infty\text{)} \quad : \quad \sum_{\nu_1} \mathbf{A}^{eff}_{\mu_1\nu_1}(t^{(1)}_{\kappa_1}, \omega) E_{\nu_1} = \omega E_{\mu_1}$$

$$\text{CIS(D)} \quad : \quad \omega^{\mathrm{CIS(D)}} = \sum_{\mu_1\nu_1} E^{\mathrm{CIS}}_{\mu_1} \mathbf{A}^{eff}_{\mu_1\nu_1}(t^{(1)}_{\kappa_1}, \omega) E^{\mathrm{CIS}}_{\nu_1}$$

The attentive reader has probably observed that the partitioned, doubles amplitude-direct formulation for second order methods—although it removes the need to store complete sets of any doubles amplitudes—does alone not reduce much the storage requirements of these methods: the calculation of the doubles amplitudes requires the electron repulsion integrals (ERIs) in the (modified) MO basis, which are obtained through four-index transformations from the AO integrals, as e.g. in Eqs. (93) and (99). Efficient implementations of such transformations require the storage of an array with half-transformed integrals of the size of $\frac{1}{2}O^2N^2$, where O is the number of occupied and N the number of atomic orbitals, which is even slightly more than needed for the doubles amplitudes. For CC2 and also for the other second order methods for excited states and in the calculation of gradients for the MP2 energies, the doubles amplitudes need to be contracted in addition with two-electron integrals with three occupied or virtual indices, $(ai|jk)$ and $(ai|bc)$, which within the schemes sketched above would give rise to even larger storage requirements. The problem can be solved with the resolution-of-the-identity approximation for electron repulsion integrals.

7 The Resolution-of-the-Identity Approximation for ERIs

The main idea behind the resolution-of-the-identity approximation[51–58] for electron repulsion integrals can be sketched as follows: With increasing atomic orbital basis sets the products of AOs appearing for the electrons 1 and 2 in the expression for the four-index two-electron integrals,

$$(\alpha\beta|\gamma\delta) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \chi_\alpha(\vec{r}_1)\chi_\beta(\vec{r}_1)\frac{1}{r_{12}}\chi_\gamma(\vec{r}_2)\chi_\delta(\vec{r}_2)d\tau_1 d\tau_2 \;, \tag{100}$$

will soon become (numerically) highly linear dependent and thus it should be possible to expand these products which good accuracy in a basis set of auxiliary functions Q,

$$\chi_\alpha(\vec{r}_1)\chi_\beta(\vec{r}_1) \approx \sum_Q Q(\vec{r}_1) c_{Q,\alpha\beta} \tag{101}$$

with a dimension much smaller then that of the original product space, $N(N+1)/2$, as illustrated in Fig. 1 for an atom with only s-type functions. The coefficients $c_{Q,\alpha\beta}$ can be determined through a least square procedure. Defining the remaining error in the expansion of an orbital pair

$$R_{\alpha\beta}(\vec{r}_1) = \chi_\alpha(\vec{r}_1)\chi_\beta(\vec{r}_1) - \sum_Q Q(\vec{r}_1) c_{Q,\alpha\beta} \;, \tag{102}$$

the quadratic error in the coulomb repulsion integrals $(\alpha\beta|\gamma\delta)$ can be written as

$$(R_{\alpha\beta}|R_{\gamma\delta}) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} R_{\alpha\beta}(\vec{r}_1)\frac{1}{r_{12}}R_{\gamma\delta}(\vec{r}_2)d\tau_1 d\tau_2 \tag{103}$$

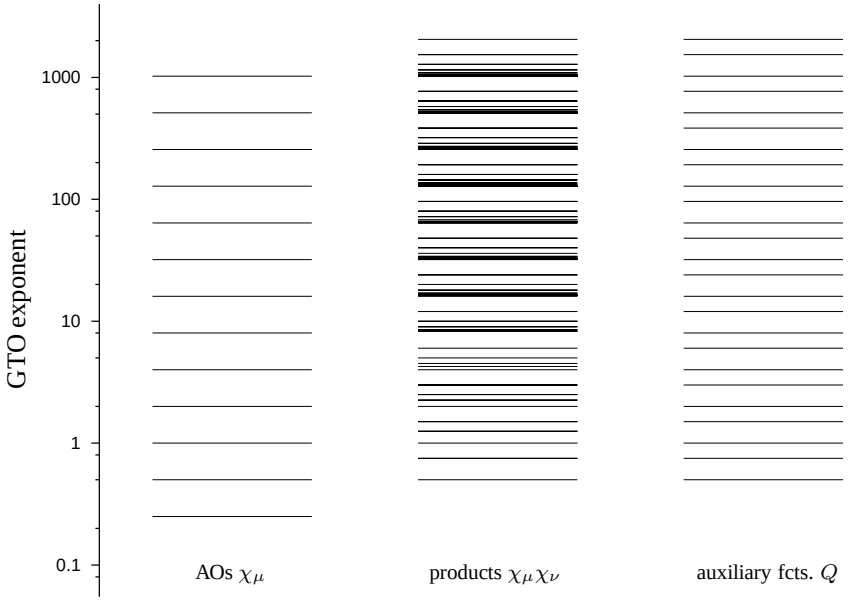


Figure 1. The left column shows exponents α_μ of an even-tempered (13s) atomic Gaussian type orbital (GTO) basis $\chi_\mu(r) = \exp(-r^2\alpha_\mu)$ and the column in the middle the exponents of all 169 overlap Gaussian functions resulting on the same atom from the products $\chi_\mu\chi_\nu$. The right column shows the exponents of an even-tempered (25s) auxiliary basis $Q(r) = \exp(-r^2\alpha_Q)$ set which could be used to expand these products.

and fulfill the Schwartz inequality

$$(R_{\alpha\beta}|R_{\gamma\delta}) \leq \sqrt{(R_{\alpha\beta}|R_{\alpha\beta})}\sqrt{(R_{\gamma\delta}|R_{\gamma\delta})}\,. \tag{104}$$

Minimization of $(R_{\alpha\beta}|R_{\alpha\beta})$ with respect to the expansion coefficients c leads to the linear equation:

$$\frac{d}{dc_{Q,\alpha\beta}}(R_{\alpha\beta}|R_{\alpha\beta}) = 0 \quad \Leftrightarrow \quad (R_{\alpha\beta}|Q) = 0 \quad \Leftrightarrow \quad (\alpha\beta|Q) - \sum_P c_{P,\alpha\beta}(P|Q) = 0 \tag{105}$$

with

$$(P|Q) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} P(\vec{r}_1)\frac{1}{r_{12}}Q(\vec{r}_2)d\tau_1 d\tau_2\,, \tag{106}$$

$$(\alpha\beta|Q) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \chi_\alpha(\vec{r}_1)\chi_\beta(\vec{r}_1)\frac{1}{r_{12}}Q(\vec{r}_2)d\tau_1 d\tau_2\,. \tag{107}$$

Arranging the two-center integrals in a matrix $V_{PQ} = (P|Q)$ the expansion coefficients can be expressed as

$$c_{Q,\alpha\beta} = \sum_P (\alpha\beta|P)[V^{-1}]_{PQ}\,, \tag{108}$$

and one obtains for the four-index coulomb integrals the approximation

$$(\alpha\beta|\gamma\delta) \approx \sum_{QP} (\alpha\beta|Q)[V^{-1}]_{QP}(P|\gamma\delta) \; . \tag{109}$$

We have above derived Eq. (109) as result of a least square fitting procedure for the overlap densities $\chi_\alpha(\vec{r})\chi_\beta(\vec{r})$, which is why this approximation is also known as "density fitting"[19,20]. Eq. (109) can be compared with the expression for an (approximate) resolution of the identity for square integrable functions in three-dimensional space,

$$\mathbf{1} \approx \sum_{QP} |Q\rangle [S^{-1}]_{QP} \langle P| \qquad \text{with} \quad S_{PQ} = \int_{\mathbb{R}^3} Q(\vec{r})P(\vec{r})d\tau \; , \tag{110}$$

applied to four-center overlap integrals

$$\int_{\mathbb{R}^3} \chi_\alpha(\vec{r})\chi_\beta(\vec{r})\chi_\gamma(\vec{r})\chi_\delta(\vec{r})d\tau = \langle \alpha\beta|\delta\gamma\rangle \approx \sum_{QP} \langle \alpha\beta|Q\rangle [S^{-1}]_{QP} \langle P|\delta\gamma\rangle \; . \tag{111}$$

We see that Eq. (109) can alternatively be viewed as an (approximate) resolution of the identity in a Hilbert space where the coulomb operator $1/r_{12}$ is used to define the scalar product as in Eqs. (100) and (103). This approximation has thus all properties expected from a resolution-of-the-identity or basis set approximation as e.g. that the norm of the error in the expansion $||R_{\alpha\beta}|| = (R_{\alpha\beta}|R_{\alpha\beta})$ will always decrease with an extension of the auxiliary basis and that the approximation becomes exact in the limit of a complete auxiliary basis set $\{Q\}$.

It is important to note that the resolution-of-the-identity approximation does not—or at least not in general—reduce the computational costs for the calculation of AO four-index electron repulsion integrals, since the right hand side of Eq. (109) is more complicated to evaluate than the left hand side. A reduction of the computational costs is only achieved if the decomposition of the four-index integrals into three- and two-index intermediates, provided by this approximation, can be exploited to simplify contractions of the AO coulomb integrals with other intermediates.

A common bottleneck of all second order correlation methods (for ground and excited states) is the four-index transformation of the AO ERIs $(\alpha\beta|\gamma\delta)$ to ERIs in a molecular orbital basis (possibly modified as in Eq. (93) or (99)) with two occupied and two virtual indices:

$$(ai|bj) = \sum_\alpha C_{\alpha a} \sum_\gamma C_{\gamma b} \sum_\beta C_{\beta i} \sum_\delta C_{\delta j} \; (\alpha\beta|\gamma\delta) \; . \tag{112}$$

Efficient algorithms for this transformation require a number of floating point multiplications that scales for the individual partial transformations with $\frac{1}{2}ON^4 + \frac{1}{2}O^2N^3 + \frac{1}{2}O^2VN^2 + \frac{1}{2}O^2V^2N$ (ignoring possible sparsities in the integrals or coefficients) and, as already pointed out above, disc space in the order of $\frac{1}{2}O^2N^2$.

Using the resolution-of-the-identity approximation, the four-index integrals in the MO basis can be obtained as

$$(ai|bj) \approx \sum_P B_{P,ai} B_{P,bj} \tag{113}$$

Table 1. Comparison of elapsed wall-clock timings for RI-MP2 vs. conventional integral-direct MP2 energy calculations (# fcts. is the number of basis functions and # e^- the number of correlated electrons, T_{MP2} timings obtained with the `mpgrad` code of the TURBOMOLE package[61]).

molecule	basis	# fcts.	# e^-	T_{MP2}	T_{RI-MP2}
benzene[a]	QZVPP	522	30	28 min	24 sec
benzene[a]	aug-cc-pVTZ	756	30	3.8 h	1.2 min
$Fe(CO)_5$[a]	QZVPP	670	66	11.3 h	8.7 min
$Fe(C_5H_5)_2$[a]	QZVPP	970	66	843 h	45 min
C_{60}[a,b]	cc-pVTZ	1800	240	112 h	171 min
Calix[4]arene[b,c]	cc-pVTZ	1528	184	39.3 h	5.6 h

[a] RI-MP2 timings for `ricc2` code of the TURBOMOLE package[61]; [b] from Ref. 62;
[c] RI-MP2 timings for `rimp2` code of the TURBOMOLE package[61];

with

$$B_{P,ai} = \sum_Q [V^{-1/2}]_{PQ} \sum_\alpha C_{\alpha a} \sum_\beta C_{\beta i}(Q|\alpha\beta) \tag{114}$$

which requires only $ON^2N_x + OVNN_x + OVN_x^2 + \frac{1}{2}O^2V^2N_x$ floating point multiplications and memory or disc space in the order ONN_x. With auxiliary basis sets optimized[56,59,60] for the application in second order methods N_x is typically 2–4 $\times N_x$. Assuming that $O \ll V \approx N$ (usually given in correlated calculations), one finds that the number of floating point operations is by the RI approximation reduced by a factor of $\approx (N/O + 3)N/N_x$. With doubly polarized or correlation-consistent triple-ζ basis sets (e.g. TZVPP or cc-pVTZ) as often used with MP2 or CC2, the RI approximation typically reduces the CPU time for the calculation of the $(ai|bj)$ integrals by more than an order of magnitude. Some typical examples for MP2 calculations for the ground state correlation energy are given in Table 1. These also demonstrate how the reduction in CPU time obtained with the RI approximation increases with the size of the orbital basis set. An important point for calculations on weakly bonded (i.e. hydrogen-bridged or van der Waals) systems is that the efficiency of the integral prescreening, which is important for the performance of conventional implementations using 4-index AO ERIs, diminishes if diffuse functions are included in the basis set. For weakly bonded complexes such diffuse functions are, however, needed for an accurate description of the long range electrostatic, exchange-correlation, and dispersion interactions. As seen at the calculations for benzene with the QZVPP and the aug-cc-pVQZ basis, RI-MP2 calculations are much less sensitive to such effects: while the CPU time for the conventional MP2 calculation increases from QZVPP to aug-cc-pVQZ by more than a factor of 8, the CPU time needed for the RI-MP2 calculation increases only by a factor of 3.

However, for large scale applications at least as important is that the scaling of the storage requirements in the calculation of the integrals $(ai|bj)$ with the system size is reduced to $\mathcal{O}(ONN_x)$. In combination with the doubles amplitude-direct formulation outlined in the previous subsection, the RI approximation completely removes the need to store any intermediates larger than $\mathcal{O}(ONN_x)$ on disc or in memory. For

example the MP2 ground state energy can now be calculated using the following algorithm:

precompute $B_{Q,ai}$
do i = 1, nocc
  do j = i, nocc
    $I^{ij}_{ab} = \sum_Q B_{Q,ai} B_{Q,bj} \quad \forall \; a,b$ (matrix-matrix multiply)
    do a = 1, nvirt
      do b = 1, nvirt
        $t^{ij}_{ab} = I^{ij}_{ab}/(\epsilon_i - \epsilon_a + \epsilon_j - \epsilon_b)$
        $E_{\mathrm{MP2}} = E_{\mathrm{MP2}} + (2-\delta_{ij})\{2I^{ij}_{ab} - I^{ij}_{ba}\}t^{ij}_{ab}$
      end do
    end do
  end do
end do

The reductions are even larger for CC2 and other second order methods for excited states and for the $\mathcal{O}(\mathcal{N}^5)$-scaling steps in the calculation of MP2 gradients. It turns out that all contractions which involve other four-index integrals in the MO basis than those of $(ia|jb)$-type, needed in second order methods, can with the decomposition given by Eq. (109) reformulated such that an explicit calculation of the four-index MO integrals can be avoided.

Together with the reduction in the CPU time the elemination of the storage bottleneck opened the possibility to apply MP2 and CC2 to much larger systems as was feasible with conventional implementations based on four-index AO ERIs. Since the steep increase of the computational costs with the basis set size is reduced by the RI approximation from $\mathcal{O}(N^4)$ to $\mathcal{O}(N^2 N_x)$ it is also easier than before to carry out such calculations with accurate basis sets, as needed to exploit fully the accuracy of MP2, CC2 or the other second order methods.

At this point it becomes neccessary to ask what are the errors introduced by the RI approximation? As is obvious from the above discussion, the accuracy (but also the efficiency) of the RI approximation depends on the choice of the auxiliary basis sets. For a balanced treatment the auxiliary basis set should be optimized for the particular orbital basis used in the calculation. Firstly, because the orbital products that need to be well represented depend strongly on the orbital basis and, secondly, because the accuracy of the approximation should increase with increasing accuracy of the orbital basis to make sure that eventually a correct basis set limit will be obtained. To fully exploit the potential of the approximation it is advantageous to further "tune" the auxiliary basis set for the integrals most important in the employed electronic structure method. For second order methods these are, as shown above, $(ai|bj)$-type integrals. The auxiliary basis functions are thus used to expand products of occupied with virtual molecular orbitals:

$$\phi_a(\vec{r})\phi_i(\vec{r}) \approx \sum_Q Q(\vec{r}) c_{Q,ai} \; . \tag{115}$$

If we consider an atom, all products will be linear combinations of Gaussian type functions centered at the atom with angular momenta up to $l_{aux} = l_{orb} + l_{occ}$, where l_{orb} is

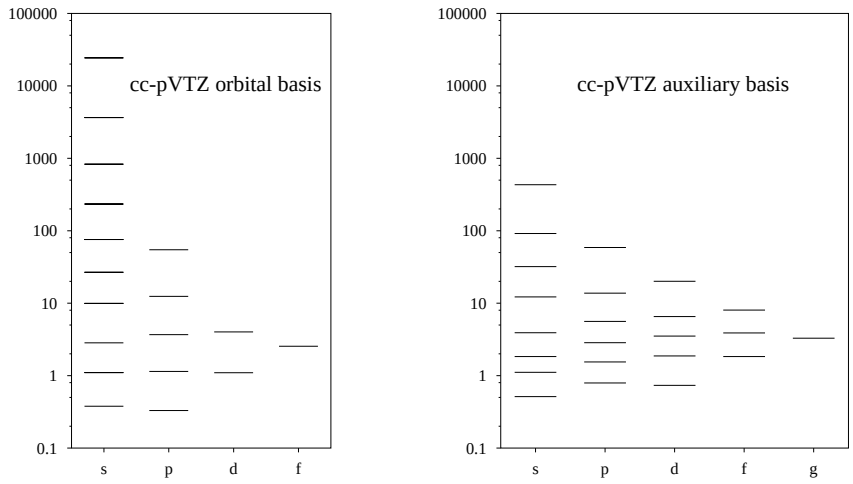


Figure 2. Exponents of the primitive GTOs in the cc-pVTZ orbital[63] (on the left) and auxiliary[59,60] (on the right) basis sets for the neon atom.

the highest angular momentum included in the orbital basis set and l_{occ} the highest angular momentum of an occupied orbital. Also the range of exponents that should be covered by the auxiliary basis can be deducted from similar considerations, but it should be taken into account that the importance of the orbital products $\phi_a\phi_i$ for electron correlation varies over orders of magnitudes. E.g., the contributions of core orbitals and similar those over very high lying tight virtual orbitals (sometimes referred to as "anti core" orbitals) is small because of large orbital energy denominators in the expression for the amplitudes. This limits the importance of tight functions in the auxiliary basis, in particular if a frozen core approximation is used and the core orbitals cannot at all contribute to the correlation treatment. In the other direction, the most diffuse exponent needed in the auxiliary basis set is bound by the exponent of any atomic orbital contributing significantly to an occupied orbital, irrespectively how diffuse functions are included in the basis set. A typical composition of an orbital basis and a respective auxiliary basis set of correlated calculations with a second order method is shown in Fig. 2 at the example of the cc-pVTZ basis sets for the neon atom.

It turns out that the above arguments, although strictly only valid for atoms, apply in practice usually also well to molecules[n]. Therefore, the auxiliary basis sets can be optimized once at the atoms for each orbital basis and then stored in a basis set library. On the TURBOMOLE web page[61] optimized auxiliary basis sets for correlated calculations with second order methods are available for several orbital basis sets including SVP[64], TZVP[65], TZVPP[56], and QZVPP[66] and most of the correlation-consistent basis sets[63,67–72] (cc-pVXZ, aug-cc-pVXZ, cc-pwCVXZ, etc.). These have been optimized[56,59,60] such that the RI error, i.e. the additional error introduced by the RI approximation, is for the

[n]An exception are the atoms with only s orbitals occupied in the ground state configuration, in particular H and Li, which in chemical bonds are often strongly polarized. For these atoms the auxiliary basis sets contain usually functions up to $l_{orb}+1$ (instead of only l_{orb}) and are often optimized on small molecules.

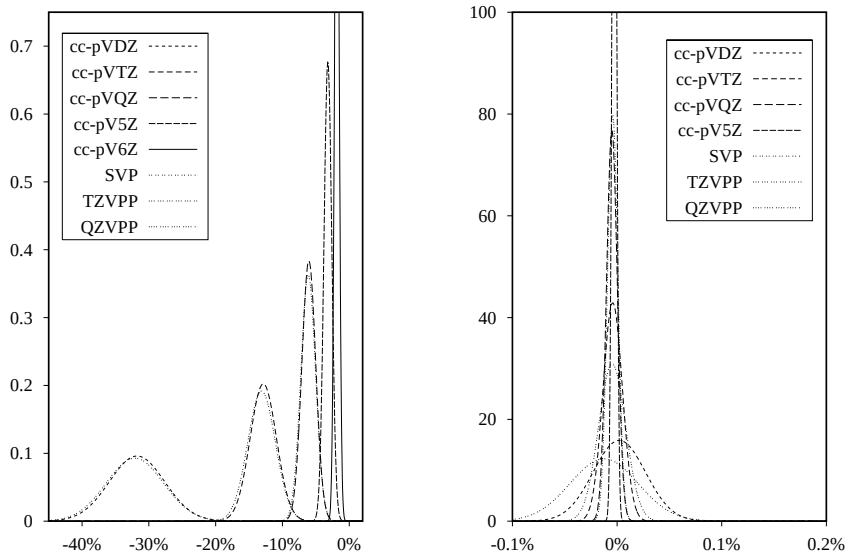


Figure 3. On the left: one-electron basis set errors in the MP2 valence correlation energy (in % of the estimated limiting value) shown as normalized Gaussian distributions determined from $\bar{\Delta}$ and Δ_{std} for a test set of 72 small and medium sized molecules with row 1 (B–Ne) and row 2 (Al–Ar) atoms[59,60]. On the right: error in the MP2 valence correlation energies due to the resolution-of-the-identity approximation for ERIs for the same test set[59,60]. Note that the scales on the abscissa differ by about three orders of magnitude!

ground state correlation energies (MP2 or CC2) about 2–3 orders of magnitudes smaller than the one-electron (orbital) basis set error of the respective orbital basis set. The correlation-consistent basis sets cc-pVXZ with X = D, T, Q, ...and the series SVP, TZVPP, QZVPP, ...constitute hierarchies that converge to the (valence) basis set limit and are thus a good example to demonstrate how orbital and auxiliary basis sets converge in parallel. Fig. 3 shows the results of an error analysis for the MP2 valence correlation energies for 72 molecules containing first and second row atoms (H, He, B–Ne, Al–Ar). The RI errors are somewhat larger for other properties than for ground state correlation energies, for which they have been optimized. In particular in response calculations for excited states the diffuse functions and also some other integral types become more important than they are for ground state calculations. But, still the RI error remains between one and two orders of magnitudes smaller than the orbital basis set error as is shown in Fig. 4 by an error analysis for RI-CC2 calculations on excited states with the aug-cc-pVTZ basis sets. Since the RI approximation is a basis set expansion approach the RI error is a smooth and usually extremely flat function of the coordinates. Therefore most of the error cancels out in the calculation of energy differences, as e.g. reaction enthalpies, and the errors in geometries are very small—typically a few 10^{-3} pm and, thus, usually below the convergence thresholds applied in geometry optimizations.

In summary, the major advantages of the resolution-of-the-identity approximation for the electron repulsion integrals for correlated second order methods are

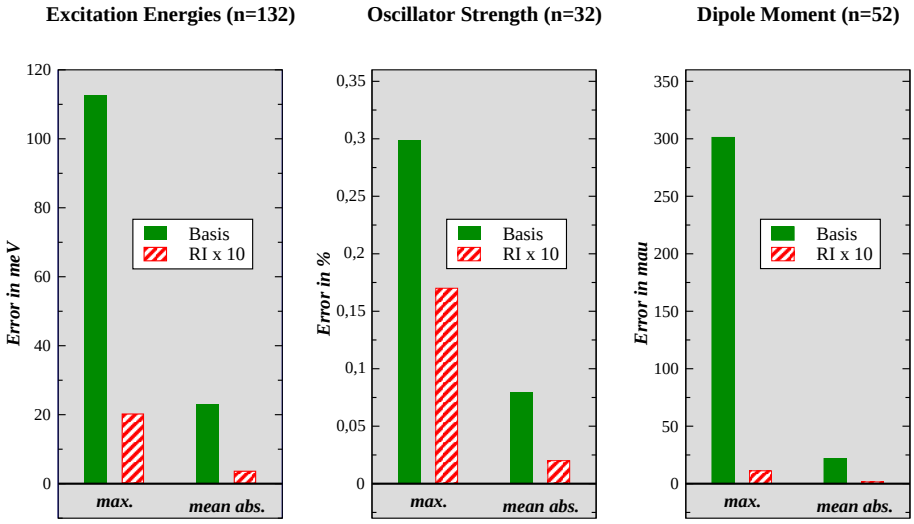


Figure 4. Mean and maximum of the one-electron orbital and the RI errors in RI-CC2 calculations for excited states with the aug-cc-pVTZ basis sets[63,69,59]. On the left: errors in excitation energies for 132 states. In the middle: errors in the oscillator strengths for 32 states. On the right: errors in the dipole moments of 52 excited states. For the test sets used and the technical details see Ref.[73], from where the data has been taken.

- It allows efficient doubles amplitude-direct implementations and eliminates the need to store any $\mathcal{O}(\mathcal{N}^4)$ arrays in memory or on disc.
- The CPU time for the correlation treatment is reduced by about an order of magnitude and more.
- It is applicable in response calculations for excited states since it does not depend on the locality of any intermediates.

Another important point related to the elimination of the huge storage demands for $\mathcal{O}(\mathcal{N}^4)$ scaling intermediates (i.e. two-electron integrals or amplitudes) is that the parallelizability of these methods is improved since less data needs to be communicated between computer nodes participating in a parallel calculation. We will come back to this point in the next section.

8 Parallel Implementation of RI-MP2 and RI-CC2 for Distributed Memory Architectures

As discussed above, the time-determining steps in RI-MP2 and other second order methods implemented with the RI approximation are the computation of the electron repulsion integrals in the MO basis $(ia|jb)$ and/or the double excitation amplitudes t_{ab}^{ij} and their contraction with integrals or other amplitudes to new intermediates, as for example

$$Y_{Q,ai} = \sum_{bj} t_{ab}^{ij} B_{Q,bj} \; . \tag{116}$$

Also for this step the computational costs increase as $\mathcal{O}(O^2V^2N_x)$. As described in Refs. 55, 73–76, $Y_{Q,ai}$ and all other intermediates calculated from t_{ab}^{ij} can efficiently be calculated in a loop over two indices for occupied orbitals with $\mathcal{O}(\mathcal{N}^2)$ memory demands. The time-determining steps of RI-MP2 can thus efficiently parallelized over pairs of indices for occupied orbitals since these are common to all steps scaling with $\mathcal{O}(O^2V^2N_x)$ or $\mathcal{O}(O^2V^3)$. An alternative could be pairs of virtual orbitals, but this would result in short loop lengths and diminished efficiency for medium sized molecules. A parallelization over auxiliary basis functions would require the communication of 4-index MO integrals between computer nodes, which would require high-performance networks. Such a solution would restrict the applicability of the program to high-end supercomputer architectures. TURBOMOLE, however, has been designed for low-cost PC clusters with standard networks (e.g. Fast Ethernet or Gigabit). Therefore we choose for the `ricc2` code a parallelization over pairs of occupied orbitals and accepted that this results in an implementation which will not be suited for massively parallel systems, since a good load balance between the participating CPUs will only be achieved for $O \gg n_{CPU}$ (vide infra).

A key problem for the parallelization of RI-MP2 and RI-CC2 is with this strategy the distribution of pairs of occupied orbitals (ij) over distributed memory nodes such that

a) the symmetry of $(ia|jb)$ with respect to permutation of $ia \leftrightarrow jb$ can still be exploited

b) the demands on the individual computer nodes for accessing and/or storing the three-index intermediates $B_{Q,ai}$ and $Y_{Q,ai}$ are as low as possible.

To achieve this, we partition the occupied orbitals into n_{CPU} batches $\mathcal{I}_m$ of (as much as possible) equal size, where n_{CPU} is the number of computer nodes. The pairs of batches $(\mathcal{I}_m, \mathcal{I}_{m'})$ with $m \leq m'$ can be ordered either on the upper triangle of a symmetric matrix or on block diagonal stripes as shown in Fig. 5. Now, each computer node gets assigned in a suitable way one block from of each diagonal, such that each node needs only access a minimal number of batches $\mathcal{I}_m$ of $B_{Q,ai}$ and $Y_{Q,ai}$. The minimal number of batches a node needs to access—in the following denoted as n_{blk}—increases approximately with $\sqrt{n_{CPU}}$. The calculation of these three-index ERIs $B_{Q,ai}$ would require about $\mathcal{O}(N^2N_X) + \mathcal{O}(ON^2N_X) \times n_{blk}/n_{CPU}$ floating point multiplications. Similar computational costs arise for some steps that involve $Y_{Q,ai}$ and other intermediates that follow the $\mathcal{O}(O^2N^2N_x)$-scaling construction of this intermediate. Thus, a conflict between minimization of the operation count and communication arises:

- If the three-index intermediates $B_{Q,ai}$ and $Y_{Q,ai}$ are communicated between the nodes to avoid multiple integral evaluations, the communication demands per node become relatively large, $\sim NN_x \times O/\sqrt{n_{CPU}}$.

- If the communication of three-index intermediates is avoided by evaluating on each node all integrals needed, the operation count for the steps which are in RI-MP2 and RI-CC2 the next expensive ones after the $\mathcal{O}(O^2V^2N_X)$ steps decreases only with $1/\sqrt{n_{CPU}}$.

The first option requires a high bandwidth for communication while the second option can also realized with a low bandwidth, but on the expense of a less efficient parallelization. For both ways a prerequisite for a satisfactory efficiency is that the total computational

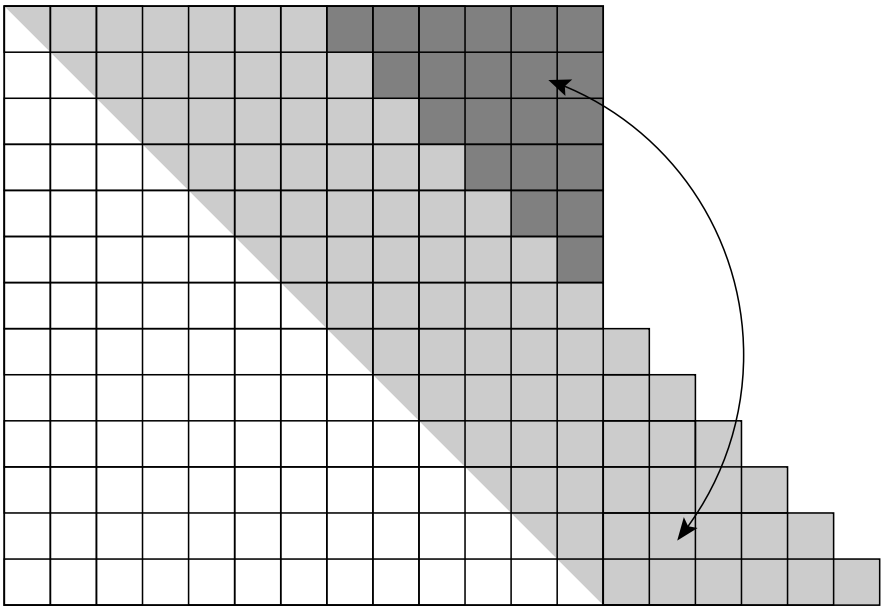


Figure 5. Arrangement of the pairs of batches $m \leq m'$ with active occupied orbitals on the upper triangle of a symmetric matrix or on block diagonal stripes.

costs are dominated by those for the $\mathcal{O}(\mathcal{N}^5)$ steps such that the time needed for multiple calculations ($\mathcal{O}(\mathcal{N}^4)$) or communication ($\mathcal{O}(\mathcal{N}^3)$) of three-index intermediates is a negligible fraction of the total time for the calculation. Both options have been realized in our parallel implementation of the `ricc2` code and shall in the following be denoted as modes for "slow communication" and "fast communication".

To implement the blocked distribution of occupied orbital indices and index pairs sketched above we define at the beginning of the calculation the following index sets:

- $\mathcal{I}_m$: a block of occupied orbitals i assigned to node m
- $\mathcal{J}_m$: merged set of the n_{blk} blocks $\mathcal{I}_n$ for which node m needs the three-index ERIs $B_{Q,ai}$ or calculates a contribution to $Y_{Q,ai}$
- $\mathcal{S}_m$: the set of all columns in the blocked distribution to which node m calculates contributions.
- $\mathcal{R}_m(n)$: the indices of the rows in column n assigned in this distribution to node m

With this concept one obtains an efficient parallelization of most program parts that involve at least one occupied index. These parts use only three- and two-index AO integrals and include all steps that scale with $\mathcal{O}(\mathcal{N}^4)$ or $\mathcal{O}(\mathcal{N}^5)$ in RI-MP2 single point calculations for energies or RI-CC2 calculations for excitation energies and spectra. For a discussion of additional demanding steps in the computation of analytic derivatives (gradients) the interested reader is referred to Refs. 55, 75–77. Here, we only sketch how the computation

of the intermediate $Y_{Q,ai}$ can be implemented without any MPI communication once each computer node has calculated or received all integral intermediates $B_{Q,ai}$ needed there:

loop $n \in \mathcal{S}_m$, loop I (where $I \subseteq \mathcal{I}_n$)

read $B_{Q,ai}$ for all $i \in I$

loop $n' \in \mathcal{R}_m(n)$, loop $j \in \mathcal{I}_{n'}$ with $j \leq i$

* read $B_{Q,bj}$
* $t_{ab}^{ij} \leftarrow B_{Q,ai} B_{Q,bj} / \left\{ \epsilon_i - \epsilon_a + \epsilon_j - \epsilon_b \right\}$
* $Y_{P,ai} \leftarrow (2t_{ab}^{ij} - t_{ba}^{ij})\ B_{P,bj}$ and for $j \neq i$ also $Y_{P,bj} \leftarrow (2t_{ab}^{ij} - t_{ba}^{ij})\ B_{P,ai}$

end loop j, loop n'

store $Y_{P,ai}$ and $Y_{P,bj}$ on disk (distributed)

end loop I, loop n

If only the RI-MP2 energy is needed, it can be evaluated directly after the calculation of the integrals $(ia|jb)$ and amplitudes t_{ab}^{ij} as described in Sec. 6.1 and the calculation of $Y_{Q,ai}$ can be skipped. If the latter intermediates are needed, the contributions to the $Y_{Q,ai}$ intermediate can be added and redistributed (after the loop over n has been closed) such that each node has the complete results for $Y_{P,ai}$ for all $i \in \mathcal{J}_m$ (requiring the communication of $\approx 2OVN_x/\sqrt{n_{CPU}}$ floating point numbers per node).

8.1 Performance for parallel RI-MP2 energy calculations

To benchmark the calculation of MP2 energies we used four typical test systems with structures as shown in Fig. 6:

- A calicheamicine model taken from Ref. 78, which has also no point group symmetry. These calculations have been done in the cc-pVTZ basis sets[63,67,68] with 934 orbital and 2429 auxiliary functions and 124 electrons have been correlated.

- The fullerene C_{60}, which has I_h symmetry, but the calculations reported here exploited only the Abelian subgroup D_{2h}. The cc-pVTZ basis set has been used, which in this case comprises 1800 orbital and 4860 auxiliary basis functions and the 240 valence electrons were correlated.

- A chlorophyll derivative which has also no point group symmetry. The cc-pVDZ basis with in total 918 orbital and 3436 auxiliary functions have been used and 264 electrons have been correlated.

- A cluster of 40 water molecules as an example for a system where integral prescreening leads to large reductions in the costs in conventional MP2 calculations. The basis sets are 6-31G$*$ for the orbital[79] and cc-pVDZ for the auxiliary[59] basis with, respectively, 760 and 3840 functions; the point group is C_1 and the 320 valence electrons have been correlated.

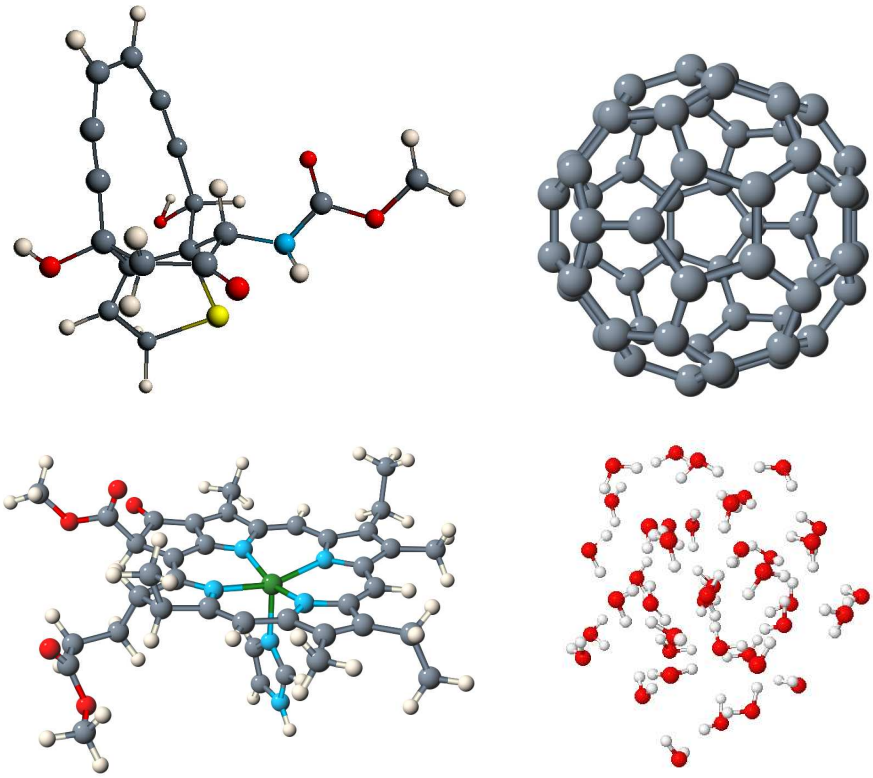


Figure 6. Structures of the four test examples used to benchmark the performance of parallel RI-MP2 calculations. For the details of the basis sets and the number of correlated electrons see text.

The maximum amount of core memory used by the program was in all calculations limited to 750 Mb. The calculations were run on two different Linux cluster: one cluster with ca. 100 Xeon Dual 2.8 GHz nodes connected through a cascaded Gigabit network and a second cluster with ca. 64 Athlon 1800MP MHz nodes connected through a 100 MBit fast Ethernet network. Due to a much larger load on the first cluster and its network the transfer rates reached in the benchmark calculations varied between ca. 80–200 MBit/sec per node. On the Athlon Cluster with the 100 MBit network we reached transfer rates of ca. 20–50 MBit/sec per node.

Fig. 7 shows timings for the calculation of MP2 energies for the C_{60} fullerene. On both architectures in sequential runs about 55% of the time are spend in the matrix multiplication for the $\mathcal{N}^5$ step. With increasing number of nodes this ratio slowly decreases. In case of the "slow communication" mode because the costs for the integral evaluation take an increasing fraction of the total wall time; in the "fast communication" mode (and here in particular on the cluster with the slower network) because of the increasing fraction of time spent in the communication of the 3-index MO integral intermediate $B_{Q,ai}$. Not parallelized steps—as e.g. the evaluation of the matrix V_{PQ} of 2-index ERIs, its Cholesky decomposition and formation of the inverse— take only a marginal fraction of the total

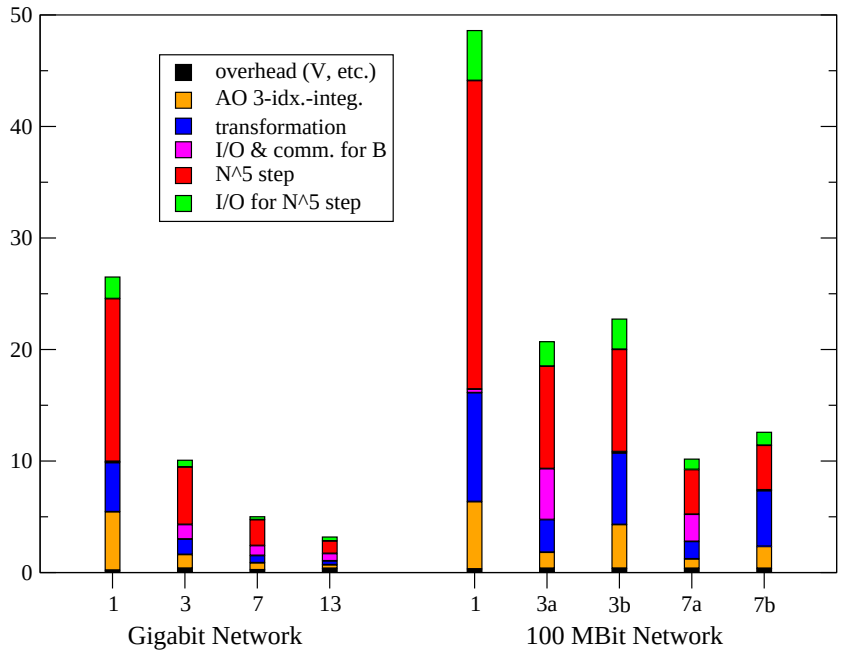


Figure 7. Timings for the most important steps in parallel RI-MP2 energy calculations for C_{60} in the cc-pVTZ basis (240 electrons correlated). For technical details of the machines used see text. At the abscissa we indicated the number of CPUs used for the calculations. For the cluster with a 100 MBit Network letters "a" and "b" are added, respectively, for calculations in the "fast" and "slow" communication modes. On the other cluster only the former program mode has been tested. The fraction denoted "overhead" includes most non-parallel steps, as the calculation of the Coulomb metric V and the inverse of its Cholesky decomposition, I/O and communication of MO coefficients, etc. With "AO 3-idx.-integ" we denoted the time spend for the calculation of the AO 3-index integrals $(P|\mu\nu)$ and with "transformation" and "I/O & comm. for B" the fractions spend in the three-index transformations for the intermediates B^i_{Qa} and for saving these intermediates on disk and/or distributing them to other computer nodes. "N^5 step" and "I/O for N^5 step" are the fractions spend, respectively, in the $\mathcal{N}^5$-scaling matrix multiplication and the I/O of B intermediates during the calculation of two-electron MO integrals. For parallel calculations idle times caused by non-perfect load-balance are included under the point "I/O for N^5 step".

wall time and the fraction of the time spend in the I/O stays approximately constant with the number of nodes used for the calculation. Another important message from Fig. 7 is, that even with a relatively slow network it is advantageous to communicate the 3-index intermediates, although on the cluster with the slower network the difference in performance between the two modes is not large. We note, however, that this depends also on the size of the system and the basis sets.

Because of the symmetry of the molecule, an RI-MP2 energy calculation for C_{60} is today not really a large scale application. The same holds for the other three test examples. Nevertheless, already for these (for parallel calculations) small examples the speed ups obtained with the present implementation are reasonable as Fig. 8 shows. The speed up obtained increases with the system size as the computational costs become dominated by the $\mathcal{N}^5$-scaling matrix multiplication in the construction of the MO 4-index ERIs and

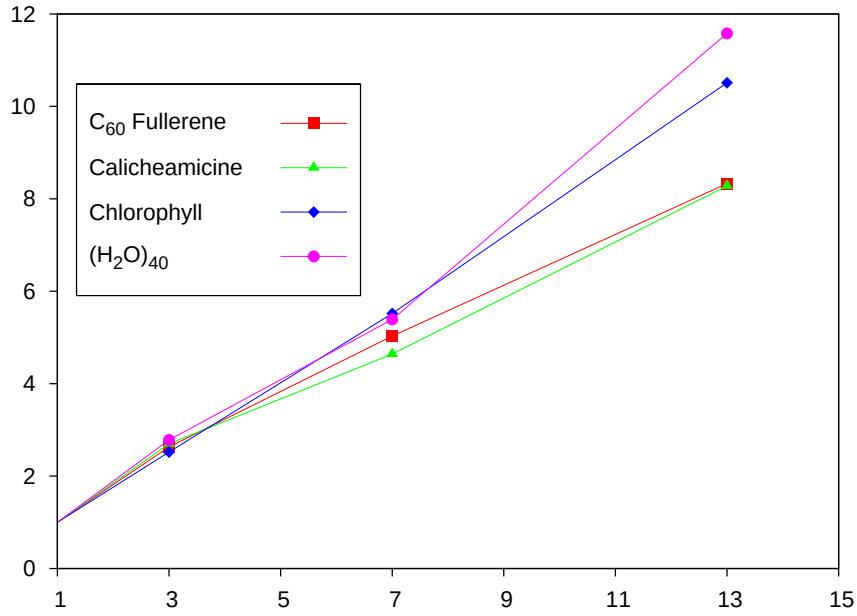


Figure 8. Speed up obtained for parallel RI-MP2 energy calculations on the Linux cluster with Gigabit network with four test examples. The number of nodes is given on the abscissa and the speed up (defined as wall time of parallel calculation dived by the wall time of the sequential run) is indicated on the ordinate.

the less good parallelizing calculation and/or communication of the 3-index MO integrals becomes unimportant for the total wall time.

9 RI-MP2 Calculations for the Fullerenes C_{60} and C_{240}

An important aspect of the parallel implementation of RI-MP2 is that it allows to combine the fast RI-MP2 approach with *parallel* Hartree-Fock self-consistent field (HF-SCF) calculations, available today in many program packages for electronic structure calculations, to optimize geometries for relatively large molecules at the MP2 level. An example for such a calculation is the determination of the MP2 basis set limit for the ground state equilibrium structure of C_{60}. The structure of C_{60} has been studied before at the MP2 level by Häser and Almlöf[81] in 1991, but due to the large computational costs of MP2 the calculations had to be limited to a singly polarized TZP basis set ([5s3p1d], 1140 basis functions), which is known to cover only about 75% of the correlation energy. With the parallel implementation of RI-MP2 it was now possible repeat this calculation using cc-pVTZ basis ([4s3p2d1f], 1800 basis functions), which gives typically correlation energies almost within 90% of the basis set limit, and the cc-pVQZ basis ([5s4p3d2f1g], 3300 basis functions), which usually cuts the remaining basis set errors again into half. The results for the bond lengths and the total energies are summarized in Table 2 together with the results from Ref. 81 and the available experimental data. As anticipated from the quality of the basis sets, the result for the correlation energy increases by about 15% from the MP2/TZP

Table 2. Equilibrium bond distances of C_{60}; d_{C-C} denotes the distance between adjacent C atoms in a five-ring and $d_{C=C}$ the distance between to the C-C bond shared between to six-rings. The bond distances are given in Ångstrøm (Å) and the total energies in Hartrees (H).

Method	d_{C-C}/Å	$d_{C=C}$/Å	Energy/hartree
SCF/DZP[a]	1.450	1.375	-2272.10290
SCF/TZP[a]	1.448	1.370	-2272.33262
MP2/DZP[b]	1.451	1.412	-2279.73496
MP2/TZP[b]	1.446	1.406	-2280.41073
MP2/cc-pVTZ[c]	1.443	1.404	-2281.65632
MP2/cc-pVQZ[c]	1.441	1.402	−2282.34442
exp.[d]	1.458(6)	1.401(10)	
exp.[e]	1.45	1.40	
exp.[f]	1.432(9)	1.388(5)	

[a] from Ref. 80; [b] from Ref. 81; [c] from Ref. 76, at the MP2/cc-pVTZ optimized structure the SCF energy is -2272.40406 hartree; [d] gas phase electron diffraction, Ref. 82; [e] solid state NMR, Ref. 83; [f] X-ray of $C_{60}(OsO_4)$(4-*tert*-butylpyridine)$_2$, Ref. 84;

to the MP2/cc-pVTZ calculation and again by about 6% from the cc-pVTZ to the cc-pVQZ basis. Also the changes in the bond lengths from the MP2/TZP to the MP2/cc-pVQZ level are with 0.004–0.005 Å of the same magnitudes as between the MP2/DZP and MP2/TZP calculations. But the difference between the two C–C distances remains almost unchanged, and also the comparison with the experimental data is not effected, since the error bars of the latter are with about ± 1 pm of the same order of magnitude as the basis set effects. The inclusion of core correlation effects would lead to a further slight contraction of the bond lengths, but the largest uncertainty comes from higher-order correlation effects which would probably increase the bond lengths in this system, but likely not more than 0.005 Å. Therefore, it is estimated that the MP2/cc-pVQZ results for the equilibrium bond distances (r_e) of the buckminster fullerene C_{60} are accurate within $\pm$ 0.005 Å. This is slightly less than the uncertainty of the presently available experimental data. Within their uncertainties the ab initio calculations and the experiments are thus in good agreement.

Another example demonstrating which system sizes can be handled with the parallel implementation of RI-MP2 is the next larger icosahedral homologue of the Buckminster fullerene C_{60}: the C_{240} molecule.The correlation consistent triple-ζ basis cc-pVTZ comprises for this molecules 7200 basis functions and, if the $1s$ core orbitals are kept frozen, 960 electrons have to be correlated. This calculation has been run on a Linux cluster with Dual Xeon 2.8 GHz nodes connected by a Gigabit network. Because the memory demands of implementation increase for non-Abelian point groups with the square of the dimension of the irreducible representations the calculation was carried out in the D_{2h} subgroup of the molecular point group I_h. On 19 CPUs the RI-MP2 calculation was completed after 16 hours and 6 minutes. About 12.5% of the time was spend in the evaluation and distribution of the two- and three-index integrals and 85% in the $\mathcal{O}(O^2V^2N_x)$ scaling construction of

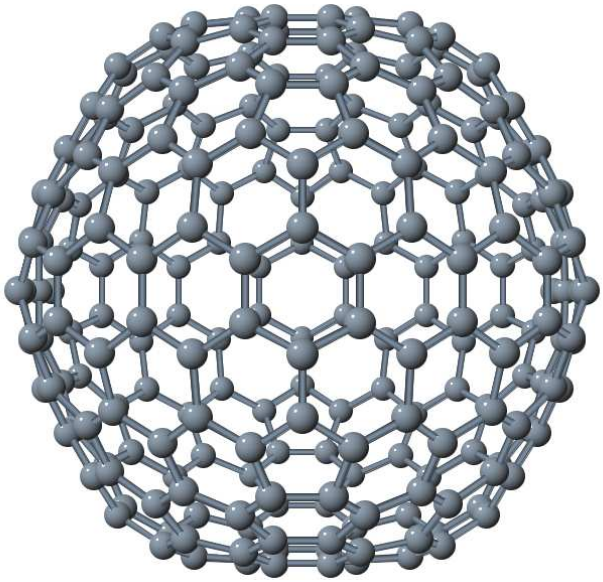


Figure 9. Structure of the icosahedral fullerene C_{240}.

the four-index integrals in the MO basis $(ia|jb)$. In D_{2h} symmetry about 6×10^{11} four-index MO integrals ($\approx$ 4.8 TByte) had to be evaluated to calculate the MP2 energy. This shows that such a calculation would with a conventional (non-RI) MP2 require either an enormous amount of disc space or many costly re-evaluations of the four-index AO two-electron integrals and would thus even on a massively parallel architecture difficult to carry out. To the best of our knowledge this is the largest canonical MP2 calculation done until today. With the parallel implementation of the RI-MP2 approach calculations of this size can now be carried out on PC clusters build with standard (and thus low cost) hardware and are expected to become soon routine applications.

The total energy of C_{240} obtained with MP2/cc-pVTZ at the BP86[85–87]/SVP[64,58] optimized structure[88] is -9128.832558 H. For the buckminster fullerene C_{60} a single point MP2/cc-pVTZ calculation at the BP86/SVP optimized geometry gives a total energy of -2281.645107 H. Neglecting differential zero-point energy effects, which in this case are expected to be small, we obtain from our calculations an estimate for the reaction enthalpy of $4 \times C_{60} \rightarrow C_{240}$ of -2.25 H, i.e. a change in the enthalpy of formation per carbon atom of -9.4 mH or -25 kJ/mol. This can be compared with the experimental result[89] for $\Delta_f H^0$ of C_{60} relative to graphite of 39.25$\pm$0.25 kJ/mol. Thus, the present calculations predict that the strain energy per carbon atom in C_{240} is with $\approx$ 15 kJ/mol only about 35% of the respective value in C_{60}.

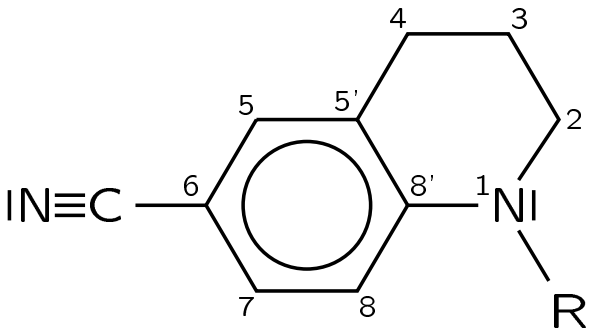


Figure 10. Enumeration of the atoms in NMC6 (R = methyl) and NTC6 (R = *tert*butyl). For DMABN R = methyl and aliphatic six-ring is replaced by a (second) methyl group at the N-atom.

10 Geometry Optimizations for Excited States with RI-CC2: The Intramolecular Charge Transfer States in Aminobenzonitrile Derivatives

An example for the optimization of excited state equilibrium structures with RI-CC2 are recent investigations[90,91] on N-alkyl-substituted aminobenzonitriles (see Fig. 10). A problem discussed for this class of molecules in the literature since several decades in many publications has been the structure of a so-called intramolecular charge-transfer (ICT) state which is observed in fluorescence and femtosecond spectroscopic experiments close to a so-called locally excited (LE) state.[92–96] The two states belong to the two lowest singlet hypersurfaces S1 and S2, which are connected through a conical intersection seam. Experimental and theoretical results[97–101] indicate that the reaction coordinate which connects the minima on the two surfaces through the conical intersection involves a Kekulé-like distortion of the phenyl ring and a twist of the amino group, which for the N,N-dimethylaminobenzonitrile (DMABN) is known to be in the ground state almost coplanar with the phenyl ring. That the twisting coordinate is involved probably explains distinct effects of different aliphatic substituents at the amino group on the fluorescence properties (vide infra) which are intensively discussed in the literature. In 1-*tert*-butyl-6-cyano-1,2,3,4-tetrahydroquinoline (NTC6) and 1-metyl-6-cyano-1,2,3,4-tetrahydroquinoline (NMC6) a twist of the amino group is restricted by the aliphatic ring to a certain range of torsion angles, but on the other side the sterically demanding bulky *tert*-butyl substituent in NTC6 disfavors a coplanar orientation. CC2/TZVPP calculations[91] predict for the ground state of NMC6 an almost coplanar orientation of the phenyl and amino moieties, but for NTC6 a tilted geometry with a twist angle of about 28° (cmp. Table 3).

Table 4 gives an overview on the CC2/TZVPP results for some spectroscopic properties of DMABN, NMC6 and NTC6, e.g. the absorption and emission energies and the dipole moments in comparison with the available experimental data. For the ICT states we found for NMC6 and NTC6 three conformations. Table 5 summarizes the results for the energetically lowest-lying structures and the ones with the highest dipole moments denoted as, respectively, ICT-1 and ICT-2, in comparison with the structure of the single conformer in the ICT state of DMABN. In all three molecules the ICT equilibrium geometries display

Table 3. Calculated bond lengths (pm) and angles (°) of the ground states of DMABN, NMC6, and NTC6 in comparison (from Ref. 91, for the enumeration of the atoms see Fig. 10).

	DMABN	NMC6	NTC6
d(C_{Ph}-N_1)[a]	137.7	138.1	139.0
d($C_8C_{8'}$)	141.4	141.2	141.2
d($C_{8'}C_{5'}$)	141.4	141.9	141.1
d(C_7C_8)	138.7	138.7	138.9
d($C_5C_{5'}$)	138.7	138.9	138.6
d(C_6C_7)	140.2	140.0	139.9
d(C_5C_6)	140.2	140.2	140.3
d(C_6C_{CN})	142.7	142.6	142.6
d(CN)	118.2	118.1	118.1
τ[b]	0	0.1	27.9
ϕ_1[c]	23	24.8	18.9
ϕ_2[d]	< 1	1	1.5

[a] bond distance between phenyl ring and amino group. [b] torsion angle, defined as dihedral angle of the normals defined by the planes C_8-$C_{8'}$-$C_{5'}$ and C_2-N_1-C_R and the bond $C_{8'}$-N_1. [c] out-of-plane angle of the bond $C_{8'}$-N_1 with respect to the plane C_2-N_1-C_R ("wagging" angle). [d] out-of-plane angle of the bond $C_{8'}$-N_1 with respect to the plane C_8-$C_{8'}$-C_5.

marked quinoid distortions of the aromatic ring system. An important finding, which was not anticipated from the experimental data that has been available in the literature, is that the aromatic ring is no longer confined to planarity in the excited state. Rather, the carbon atom labeled 8' in Fig. 10 is pyramidalized. Therefore the aliphatic six-ring can accommodate twist angles of the amino group of up to 60–70°, as illustrated in Fig. 11, and in this way energetically low-lying twisted ICT states can be realized even in NTC6 and NMC6. In the literature it was before assumed that the aliphatic six-ring, which connects the amino group with the phenyl ring restricts these molecules to "planarized" structures and makes such a twist impossible.

The transition to the ICT state is at the ground state geometry dominated by the one-electron HOMO→LUMO excitation in these molecules. Both orbitals are of Ph-N antibinding character, but the orbital energy of the LUMO decreases slightly faster with increasing twisting angle than the energy of the HOMO and already such a simple model predicts for the ICT state close to the ground state geometry a gradient directed to a twisted structure. With increasing twisting angle the transition assumes an increasing contribution from the HOMO-2→LUMO excitation. The HOMO-2 is the Ph-N binding counterpart of the HOMO and increases in energy with the twisting angle and mixes with the HOMO. As the angle approaches 90° one of the two orbitals becomes the lone-pair at the amino N-atom while the other is localized in the aromatic system and the transition to the ICT state is dominated by the $n \rightarrow \pi^\star$ excitation. In a many electron picture this change in the character of the excitation corresponds to an avoided crossing of S2 with another, at

Table 4. Calculated absorption and emission energies and dipole moments for DMABN, NMC6 and NTC6 in comparison with experimental data. The CC2 results for absorption and emission energies are vertical electronic transition energies; the dipole moments were calculated as analytic derivatives of the CC2 total energies.

	DMABN		NMC6		NTC6	
	CC2[a]	exp.	CC2[a]	exp.	CC2[a]	exp.
absorption (S_1) [eV]	4.41[b]	4.25[c]	4.31[d]		4.33[d]	
absorption (S_2) [eV]	4.77[b]	4.56[c]	4.58[d]	4.32[e]	4.43[d]	4.14[e]
osc. strengths (S_1)	0.03[bf]		0.03[f]		0.03[f]	
osc. strengths (S_2)	0.62[bf]		0.49[f]		0.51[f]	
T_e (LE) [eV]	4.14		4.07		3.91	
emission (LE) [eV]	3.78[g]	3.76[h]	3.67[g]	3.67[e]	3.34[g]	3.50[e]
T_e (ICT) [eV]	4.06–4.16[i]		4.18		3.71	
emission (ICT) [eV]	2.49–3.27[ig]	2.8–3.2[j]	2.53[gk]		2.51[gk]	2.8[l]–3.3[e]
dipole (GS) [D]	7.4	6.6[j]	7.5	6.8[m]	7.7	6.8[m]
dipole (LE) [D]	10.1	9.7[j]	10.4	10.6[m]	12.6	
dipole (ICT) [D]	13.3–15.1[i]	17 ± 1[j]	12.7[k]		13.5[k]	17–19[m]

[a] Unless otherwise indicated the CC2 results for DMABN are taken from Ref. 90 and those for NMC6 and NTC6 from Ref. 91. [b] CC2/TZVPP (Ref.[102]). [c] EELS band maximum (Ref. 103). [d] Vertical excitation energy to the L_a (or S_2) state which has a significantly larger oscillator strength. [e] Experimental band maximum in n-hexane (Ref. 94). [f] Oscillator strength for vertical electronic transition calculated at the CC2/TZVPP level in length gauge. [g] Vertical energy separation from ground state at the excited state equilibrium structure. [h] Maximum of dispersed emission from jet-cooled DMABN (Ref. 104). [i] The first value is the result for the gas phase equilibrium structure and the second value is obtained at the C_{2v} symmetric saddle point (Ref. 90). [j] Emission energy from ICT state from maxima of fluorescence bands; ground state dipole moment derived from the dielectric constant and refractive index in dioxane and the excited state dipole moments from time-resolved microwave conductivity measurements in dioxane (Ref. 105). [k] Value refers to the ICT-2 conformer. [l] Experimental band maximum in methanol (Ref. 94). [m] Derived from solvatochromic shift of fluorescence maximum (Ref. 94).

the ground state structure energetically higher lying, charge-transfer state—in DMABN according to DFT/SCI calculation in Ref. 106 the S5 state. The avoided crossing with this state is the main driving force for the formation of the TICT structures (twisting and pyramidilization at the $C_{8'}$ atom) in DMABN, NTC6, NMC6 and other alkyl-substituted amino-benzonitrils. It leads to a pronounced stabilization of the ICT state at large twisting angles and enhances the charge-transfer character, as it is apparent from the expectation values for the dipole moment (see Table 4). For all three molecules, DMABN, NMC6, and NTC6, one finds a similar change in the electronic character from the vertical excitation in the Franck-Condon region to the equilibrium geometries of the ICT states. This is in line with the interpretation of recent measurements of the short-time dynamics in DMABN derivatives after excitation to S2.[97,107–110]

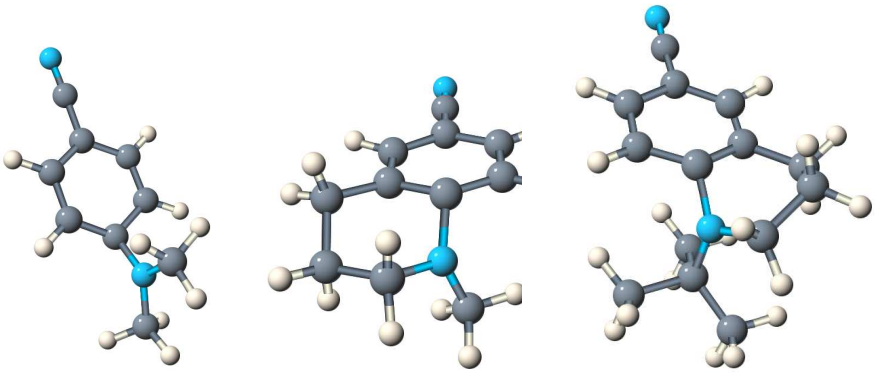


Figure 11. Equilibrium structures of the ICT states in DMABN, NMC6, and NTC6.

For NTC6 the increase in the twist angle from the ground to the excited ICT states reduces the steric strain of the *tert*-butyl group und thus compensates for the hindrance of the twist by the aliphatic bridge. We obtain at the CC2/TZVPP level that for NTC6 and DMABN the ICT states are energetically slightly below the LE state, which is reached by an one-electron transition from the PH-N antibinding HOMO to a Ph-N non-binding orbital. For NMC6, however, the inhibition of a 90° twist is not compensated by the release of a similar strain since the methyl substituent is sterically much less demanding. Thus, in difference to DMABN and NTC6 the LE→ICT reaction for NMC6 is predicted by the RI-CC2 calculations to be slightly endotherm. This explains why NMC6 is not dual fluorescent, in contrast to DMABN and NTC6.

11 Summary

The computational costs of wavefunction based correlated ab initio methods that treat the electron–electron interaction correctly through second order (so-called second order or approximate doubles methods) have in conventional implementations been dominated by the huge operation counts for the calculation of the four-index electron repulsion integrals in the AO basis and their transformation to the MO basis. The costs for these steps increase rapidly with the size of the system studied and the basis sets used. In addition, also the huge storage demands for the four-index transformation hindered applications on large systems.

With the resolution-of-the-identity approximation for the electron repulsion integrals the CPU time for the calculation of the MO integrals needed in second order methods is reduced by about an order of magnitude (and sometimes even much more) and the scaling of the storage demands is reduced from $\mathcal{O}(O^2N^2)$ to $\mathcal{O}(OVN_x)$. If optimized auxiliary basis sets are used, as they today are available for many orbital basis sets, the errors due to RI approximation are insignificant compared to the errors due to the incompleteness of the orbital basis sets.

In combination with a new parallel implementation in TURBOMOLE for distributed memory architectures (e.g. PC clusters) it became now possible to carry out RI-MP2 calculations for energies and structures with several thousands of basis functions and several

Table 5. Calculated bond lengths (pm) and angles (°) and weights of the two most important one-electron excitations (%) for the intramolecular charge-transfer states of DMABN, NMC6, and NTC6 in comparison (from Ref. 91, for the enumeration of the atoms see Fig. 10).

	DMABN	NMC6		NTC6	
	ICT	ICT-1	ICT-2	ICT-1	ICT-2
d(C_{Ph}-N_1)[a]	144.3	146.8	145.0	146.8	145.7
d($C_8C_{8'}$)	144.6	143.5	144.8	142.9	144.9
d($C_{8'}C_{5'}$)	144.6	146.2	144.5	146.0	143.6
d(C_7C_8)	137.2	137.2	137.7	137.3	137.8
d($C_5C_{5'}$)	137.2	138.0	136.9	137.9	137.1
d(C_6C_7)	142.9	143.4	142.4	143.8	142.4
d(C_5C_6)	142.9	141.8	143.7	141.9	144.0
d(C_6C_{CN})	140.9	141.2	140.9	141.1	140.8
d(CN)	118.9	118.8	118.9	118.8	118.9
τ[b]	90	54.3	66.6	58.5	65.0
ϕ_1[b]	0	24.1	14.7	20.7	5.2
ϕ_2[b]	41	43.9	44.6	36.4	43.4
HOMO→LUMO		65	62	69	64
HOMO-2→LUMO		15	17	25	16

[a] bond distance between phenyl ring and amino group. [b] for the definition of the torsion and the out-of-plane angles see Table 3.

hundreds of correlated electrons. This extends the applicability of MP2 to systems which else can only be treated with SCF or DFT methods. Calculations on excited states using e.g. the approximate coupled-cluster singles and doubles method CC2 or the perturbative doubles correction to configuration interaction singles, CIS(D), are somewhat more involved and structure optimizations for excited states are (because of weakly avoided crossings or conical intersections) much less straightforward than for ground states. With the parallel implementation of RI-CC2 they become still feasible for molecules with more than 30 atoms and many hundred basis functions even if the molecular structure has no point group symmetry.

Acknowledgments

The author in indebted to A. Köhn and A. Hellweg for their contributions to the RICC2 program and to the calculations reviewed in this manuscript. Financial support by the Deutsche Forschungsgemeinschaft (DFG) for the reported work is gratefully acknowledged.

References

1. P. Pulay. *Chem. Phys. Lett.*, 100:151–154, 1983.
2. M. Schütz, G. Hetzer, and H.-J. Werner. *J. Chem. Phys.*, 111:5691–5705, 1999.
3. G. Hetzer, M. Schütz, H. Stoll, and H. J. Werner. *J. Chem. Phys.*, 113:9443–9455, 2000.
4. G. E. Scuseria and P. Y. Ayala. *J. Chem. Phys.*, 111:8330–8343, 1999.
5. P. E. Maslen and M. Head-Gordon. *J. Chem. Phys.*, 109:7093–7099, 1998.
6. S. H. Li, J. Ma, and Y. S. Jiang. *J. Comp. Chem.*, 23:237–244, 2002.
7. K. Morokuma. *Philosophical Transactions of the Royal Society of London Series A – Mathematical Physical and Engineering Sciences*, 360:1149–1164, 2002.
8. M. Schütz and H. J. Werner. *J. Chem. Phys.*, 114:661–681, 2001.
9. M. Schütz. *Phys. Chem. Chem. Phys.*, 4:3941–3947, 1993.
10. S. Sæbo and P. Pulay. *J. Chem. Phys.*, 87:3975–3983, 2001.
11. D. S. Lambrecht and C. Ochsenfeld. *J. Chem. Phys.*, 123:184102, 2005.
12. S. Grimme. *J. Chem. Phys.*, 118:9095–9102, 2003.
13. S. Grimme. *J. Comput. Chem.*, 24:1529–1537, 2003.
14. Y. S. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon. *J. Chem. Phys.*, 121:9793–9802, 2004.
15. R. C. Lochan, Y. S. Jung, and M. Head-Gordon. *J. Phys. Chem. A*, 109:7598–7605, 2005.
16. T. D. Crawford and R. A. King. *Chem. Phys. Lett.*, 366:611–622, 2002.
17. T. Korona and H. J. Werner. *J. Chem. Phys.*, 114:3006–3019, 2003.
18. T. Korona, K. Pflüger, and H. J. Werner. *Phys. Chem. Chem. Phys.*, 6:2059–2065, 2004.
19. M. Schütz and F. R. Manby. *Phys. Chem. Chem. Phys.*, 5:3349–3358, 2003.
20. H.-J. Werner, F. R. Manby, and P. J. Knowles. *J. Chem. Phys.*, 118:8149, 2003.
21. P. Pulay. *J. Chem. Phys.*, 73:393, 1980.
22. P. Pulay. *J. Comp. Chem.*, 4:556, 1982.
23. T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular Electronic-Structure Theory*. John Wiley & Sons, New York, 2000.
24. S. Grimme and M. Waletzke. *Phys. Chem. Chem. Phys.*, 2:2075–2081, 2000.
25. B. O. Roos, K. Andersson, M. P. Fülscher, P. A. Malmqvist, L. Serrano-Andres, K. Pierloot, and M. Merchan. *Advances in Quantum Chemistry*, 93:219–331, 1996.
26. J. Olsen and P. Jørgensen. *J. Chem. Phys.*, 82:3235, 1985.
27. J. Olsen and P. Jørgensen. *Time-Dependent Response Theory with Applications to Self-Consistent Field and Multiconfigurational Self-Consistent Field Wave Functions*, In D. R. Yarkony, editor, *Modern Electronic Structure Theory*, volume 2, chapter 13, pages 857–990. World Scientific, Singapore, 1995.
28. O. Christiansen, P. Jørgensen, and C. Hättig. *Int. J. Quantum Chem.*, 68:1–52, 1998.
29. C. Hättig. *Accurate Coupled Cluster Calculation of Nonlinear Optical Properties of Molecules*, In M. G. Papadopoulos, editor, *On the non-linear optical response of molecules, solids and liquids: Methods and applications*. Research Signpost, 2003.
30. P. Salek, O. Vahtras, T. Helgaker, and H. Agren. *J. Chem. Phys.*, 117:9630–9645, 2002.
31. O. Christiansen, S. Coriani, J. Gauss, C. Hättig, P. Jørgensen, F. Pawłowski, and A.

Rizzo. *Accurate NLO Properties for small molecules. Methods and results*, In M. G. Papadopoulos, J. Leszczynski, and A. J. Sadlej, editors, *Nonlinear optical properties: From molecules to condensed phases*. Kluwer Academic Publishers, London, 2006.
32. O. Christiansen, H. Koch, and P. Jørgensen. *Chem. Phys. Lett.*, 243:409–418, 1995.
33. O. Christiansen, H. Koch, P. Jørgensen, and T. Helgaker. *Chem. Phys. Lett.*, 263:530, 1996.
34. J. Paldus and X. Z. Li. *Advances in Quantum Chemistry*, 110:1–175, 1999.
35. R. J. Bartlett. *Recent Advances in Coupled-Cluster Methods*, volume 3 of *Recent Advances in Computational Chemistry*. World Scientific, Singapore, 1997.
36. R. J. Bartlett. Coupled-cluster theory: An overview of recent developments. In D. R. Yarkony, editor, *Modern Electronic Structure Theory*, pages 1047–1131, Singapore, 1995. World Scientific.
37. M. Head-Gordon, M. Oumi, and D. Maurice. *Mol. Phys.*, 96:593–574, 1999.
38. M. Head-Gordon, R. J. Rico, M. Oumi, and T. J. Lee. *Chem. Phys. Lett.*, 219:21–29, 1994.
39. J. Schirmer. *Phys. Rev. A*, 26:2395–2416, 1981.
40. A. B. Trofimov and J. Schirmer. *J. Phys. B*, 28:2299–2324, 1995.
41. E. S. Nielsen, P. Jørgensen, and J. Oddershede. *J. Chem. Phys.*, 73:6238–6246, 1980.
42. M. J. Packer, E. K. Dalskov, T. Enevoldsen, H. J. A. Jensen, and J. Oddershede. *J. Chem. Phys.*, 105:5886–5900, 1980.
43. O. Christiansen, K. L. Bak, H. Koch, and S. P. A. Sauer. *Chem. Phys. Lett.*, 284:47–55, 1998.
44. A. I. Krylov, C. D. Sherrill, E. F. C. Byrd, and M. Head-Gordon. *J. Chem. Phys.*, 109:10669–10678, 1998.
45. T. B. Pedersen, H. Koch, and C. Hättig. *J. Chem. Phys.*, 110:8318–8327, 1999.
46. K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon. *Chem. Phys. Lett.*, 157:479, 1989.
47. H. Koch, O. Christiansen, P. Jørgensen, A. Sánchez de Merás, and T. Helgaker. *J. Chem. Phys.*, 106:1808, 1997.
48. O. Christiansen, H. Koch, and P. Jørgensen. *J. Chem. Phys.*, 105:1451–1459, 1996.
49. O. Christiansen, H. Koch, and P. Jørgensen. *J. Chem. Phys.*, 103:7429–7441, 1995.
50. C. Hättig and A. Köhn. *J. Chem. Phys.*, 117:6939–6951, 2002.
51. J. L. Whitten. *J. Chem. Phys.*, 58:4496–4501, 1973.
52. B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin. *J. Chem. Phys.*, 71:3396–3402, 1979.
53. O. Vahtras, J. E. Almlöf, and M. W. Feyereisen. *Chem. Phys. Lett.*, 213:514–518, 1993.
54. M. W. Feyereisen, G. Fitzgerald, and A. Komornicki. *Chem. Phys. Lett.*, 208:359–363, 1993.
55. F. Weigend and M. Häser. *Theor. Chem. Acc.*, 97:331, 1997.
56. F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs. *Chem. Phys. Letters*, 294:143, 1998.
57. A. P. Rendell and T. J. Lee. *J. Chem. Phys.*, 101:400–408, 1994.
58. K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs. *Chem. Phys. Lett.*, 240:283–289, 1995. Erratum ibid. 242 (1995) 652.
59. F. Weigend, A. Köhn, and C. Hättig. *J. Chem. Phys.*, 116:3175, 2001.

60. C. Hättig. *Phys. Chem. Chem. Phys.*, 7:59–66, 2005.
61. TURBOMOLE, Program Package for ab initio Electronic Structure Calculations , `http://www.turbomole.com`.
62. A. Köhn and C. Hättig. *Chem. Phys. Lett.*, 358:350–353, 2002.
63. T. H. Dunning. *J. Chem. Phys.*, 90:1007–1023, 1989.
64. A. Schäfer, H. Horn, and R. Ahlrichs. *J. Chem. Phys.*, 97(4):2571–2577, 1992.
65. A. Schäfer, C. Huber, and R. Ahlrichs. *J. Chem. Phys.*, 100(8):5829–5835, 1994.
66. F. Weigend, F. Furche, and R. Ahlrichs. *J. Chem. Phys.*, 119:12753–12762, 2003.
67. R. A. Kendall, T. H. Dunning, and R. J. Harrison. *J. Chem. Phys.*, 96:6796–6806, 1992.
68. D. E. Woon and T. H. Dunning. *J. Chem. Phys.*, 98:1358–1371, 1993.
69. D. E. Woon and T. H. Dunning. *J. Chem. Phys.*, 100:2975–2988, 1994.
70. A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning. *J. Chem. Phys.*, 110:7667–7676, 1999.
71. T. H. Dunning, K. A. Peterson, and A. K. Wilson. *J. Chem. Phys.*, 114:9244–9253, 2001.
72. K. A. Peterson and T. H. Dunning. *J. Chem. Phys.*, 117:10548–10560, 2002.
73. C. Hättig and A. Köhn. *J. Chem. Phys.*, 117:6939–6951, 2002.
74. C. Hättig and F. Weigend. *J. Chem. Phys.*, 113:5154–5161, 2000.
75. C. Hättig. *J. Chem. Phys.*, 118:7751–7761, 2003.
76. C. Hättig, A. Hellweg, and A. Köhn. *Phys. Chem. Chem. Phys.*, 2005. submitted for publication.
77. A. Köhn and C. Hättig. *J. Chem. Phys.*, 119:5021–5036, 2003.
78. P. Pulay, S. Saebø, and K. Wolinski. *Chem. Phys. Lett.*, 344:543, 2001.
79. W. J. Hehre, R. Ditchfield, and J. A. Pople. *J. Chem. Phys.*, 56:2257, 1972.
80. G. E. Scuseria. *Chem. Phys. Lett.*, 176:423, 1991.
81. M. Häser, J. Almlöf, and G. E. Scuseria. *Chem. Phys. Lett.*, 181:497–500, 1991.
82. K. Hedberg, L. Hedberg, D. S. Bethune, C. A. Brown, H. C. Dorn, R. D. Johnson, and M. De Vries. *Science*, 254:410–412, 1991.
83. C. S. Yannoni, P. P. Bernier, D. S. Bethune, G. Meijer, and J. R. Salem. *J. Am. Chem. Soc.*, 113:3190, 1991.
84. J. M. Hawkins, A. Meyer, T. A. Lewis, S. Lorin, and F. J. Hollander. *Science*, 252:312, 1991.
85. J. P. Perdew. *Phys. Rev. B*, 33(12):8822–8824, 1986.
86. J. P. Perdew. *Phys. Rev. B*, 34:7046, 1986.
87. A. D. Becke. *Phys. Rev. A*, 38(6):3098–3100, 1988.
88. F. Furche, 2005. Private communication.
89. B. V. Lebedev, L. Y. Tsvetkova, and K. B. Zhogova. *Thermochimica Acta*, 299:127–131, 1997.
90. A. Köhn and C. Hättig. *J. Am. Chem. Soc.*, 126:7399–7410, 2004.
91. A. Hellweg, C. Hättig, and A. Köhn. *J. Am. Chem. Soc.*, 2005. submitted for publication.
92. W. Rettig, B. Bliss, and K. Dirnberger. *Chem. Phys. Lett.*, 305:8–14, 1999.
93. K. A. Zachariasse. *Chem. Phys. Lett.*, 320:8–13, 2000.
94. K. A. Zachariasse, S. I. Druzhinin, W. Bosch, and R. Machinek. *J. Am. Chem. Soc.*, 126:1705–1715, 2004.

95. S. Techert and K. Zachariasse. *J. Am. Chem. Soc.*, 126:5593–5600, 2004.
96. T. Yoshihara, S. Druzhinin, and K. Zachariasse. *J. Am. Chem. Soc.*, 126:8535–3539, 2004.
97. W. Fuß, K. K. Pushpa, W. Rettig, W. E. Schmid, and S. A. Trushin. *Photochem. Photobiol. Sci.*, 1:255–262, 2002.
98. S. Zilberg and Y. Haas. *J. Phys. Chem. A*, 106:1–11, 2002.
99. W. M. Kwok, C. Ma, P. Matousek, A. W. Parker, D. Phillips, W. T. Toner, M. Towrie, and S. Umapathy. *J. Phys. Chem. A*, 105:984–990, 2001.
100. J. Dreyer and A. Kummrow. *J. Am. Chem. Soc.*, 122:2577–2585, 2000.
101. D. Rappoport and F. Furche. *J. Am. Chem. Soc.*, 126:1277–1284, 2004.
102. A. Köhn. PhD thesis, Universität Karlsruhe, 2003.
103. C. Bulliard, M. Allan, G. Wirtz, E. Haselbach, K. A. Zachariasse, N. Detzer, and S. Grimme. *J. Phys. Chem. A*, 103:7766–7772, 1999.
104. U. Lommantzsch, A. Gerlach, C. Lahmann, and B. Brutschy. *J. Phys. Chem. A*, 102:6421–6435, 1998.
105. W. Schuddeboom, S. A. Jonker, J. M. Warman, U. Leinhos, W. Kühnle, and K. A. Zachariasse. *J. Phys. Chem.*, 96:10809–10819, 1992.
106. A. B. J. Parusel, G. Köhler, and S. Grimme. *J. Phys. Chem. A*, 102:6297–6306, 1998.
107. S. A. Trushin, T. Yatsuhashi, W. Fuß, and W. E. Schmid. *Chem. Phys. Lett.*, 376:282–291, 2003.
108. T. Yatsuhashi, S. A. Trushin, W. Fuß, , W. Rettig, W. E. Schmid, and S. Zilberg. *Chem. Phys.*, 296:1, 2004.
109. W. Fuß, W. Rettig, W. E. Schmid, S. A. Trushin, and T. Yatsuhashi. *Farad. Disc.*, 127:23, 2004.
110. W. Fuß, 2005. private communication.

Density Functional Theory and Linear Scaling

Rudolf Zeller

Institute for Solid State Research and Institute for Advanced Simulation
Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: ru.zeller@fz-juelich.de

The basic concepts of density functional theory and of linear-scaling techniques to solve the density functional equations are introduced. The Hohenberg-Kohn theorem, the one-to-one mapping to an auxiliary non-interacting electron system to obtain the single-particle Kohn-Sham equations, and the construction of approximations for the exchange-correlation functional are explained. The principle of nearsightedness of electronic matter and its importance to achieve linear scaling are discussed. Finally, a recently in Jülich developed linear-scaling algorithm for metallic systems is presented and its suitability for large supercell calculations is illustrated.

1 Introduction

In the last decades density functional theory has emerged as a powerful tool for the quantum mechanical description of chemical and physical properties of materials. Density functional theory is an approach to treat the many-electron problem by single-particle equations. Instead of the many-electron wavefunction, which depends on $3N$ electronic space coordinates and N spin variables (here N is the number of electrons in the considered system), the basic quantity in density functional theory is the electron density $n(\underline{r})$, which depends on only three space coordinates. This obviously represents a considerable simplification for calculating, understanding and predicting material properties. The idea to use the density instead of the many-electron wavefunction was proposed by Thomas[1] and Fermi[2] already in 1927. The idea was fundamentally justified by the theorem of Hohenberg and Kohn[3] in 1964, which states that the ground-state energy of the many-electron system is uniquely determined by the ground-state density $n_0(\underline{r})$. Modern density functional theory has motivated an enormous number of applications primarily in the electron theory of atoms, molecules and solids, but density functional theory can be used also in the physics of liquids[4] and in nuclear physics[5].

However, although density functional theory accomplishes a considerable simplification, calculations for systems with many atoms still represent a serious computational challenge even after decades of effort to develop and improve computational techniques for the solution of the density functional equations. Systems with up to a few hundred atoms can be treated routinely today, but systems with thousands of atoms require overwhelming computing effort, because the computing time increases cubically with system size. In the last decade considerable work has been done to reduce the computational effort and linear scaling techniques have emerged as an approach to treat large systems with almost similar accuracy as available in standard techniques with cubic scaling.

The plan of this lecture is to introduce the concepts of density functional theory, to explain the reasons why linear scaling should be possible, to present the ideas used in several linear scaling techniques and finally to present an algorithm for metallic systems which was recently developed in our institute.

2 Density Functional Theory

To simplify the discussion[a] the consideration will be restricted here to a non-relativistic, non-spin-polarized, time-independent many-electron system moving in a potential provided by the electrostatic Coulomb interaction with atomic nuclei assumed at fixed positions. For this system the Hamilton operator $\hat{H}$ is given by a sum of the kinetic energy and the electron-electron, electron-nuclear and nuclear-nuclear interaction terms. Under the assumption that the nuclei are fixed the many-electron Schrödinger equation for N electrons is given by

$$\hat{H}\Psi = \left[-\frac{\hbar^2}{2m} \sum_i^N \nabla_i^2 + \sum_{i<j}^N \sum U(\underline{r}_i, \underline{r}_j) + \sum_i^N v_{ext}(\underline{r}_i) \right] \Psi = E\Psi, \qquad (1)$$

where $U(\underline{r}, \underline{r}') = e^2 |\underline{r} - \underline{r}'|^{-1}$ is the electron-electron interaction and $v_{ext}(\underline{r})$ the external potential, which contains the static potential arising from the interaction of the electrons with the nuclei and a constant term arising from the nuclear-nuclear interaction. Extensions of density functional theory to non-degenerate ground states, to spin-polarized and relativistic systems, to excited states and finite temperatures, to time-dependent and to superconducting situations are possible and can be found in the literature. Here, due to limited space, a discussion of these extensions is not possible.

2.1 Hohenberg-Kohn Theorem

The formal solution of the many-electron Schrödinger equation (1) defines a mapping from the external potential to the many-electron wavefunctions and thus also a mapping from external potential to the ground state wavefunction Ψ_0 and to the ground-state density $n_0(\underline{r})$. The first part of the Hohenberg-Kohn theorem states that the mapping can be inverted so that the external potential is uniquely determined by the ground-state density except for a trivial additive constant shift of the external potential. Because of the mapping from the ground-state density to the external potential and of the mapping from the external potential to the many-electron wavefunctions, there is also a mapping from the ground-state density to the many-electron wavefunctions and to every expectation value $\langle \Psi | \hat{O} | \Psi \rangle$, which means that every quantum mechanical observable is uniquely determined as a functional[b] of the ground-state density. The second part of the Hohenberg-Kohn theorem states that the total energy functional $E[n(\underline{r})]$ is minimal, if $n(\underline{r})$ is the ground-state density $n_0(\underline{r})$, and that the minimum $E_0 = E[n_0(\underline{r})]$ is the ground-state energy.

The proof of the Hohenberg-Kohn theorem for non-degenerate ground states proceeds by reductio ad absurdum and requires two steps. First it is shown that two potentials v_{ext} and v'_{ext}, which differ by more than a trivial constant $v_{ext} \neq v'_{ext} + const$, cannot lead to the same ground-state wavefunction Ψ_0 and then it is shown that two different ground-state wavefunctions Ψ_0 and Ψ'_0 (arising from two different potentials $v_{ext} \neq v'_{ext} + const$) cannot lead to the same ground-state density $n_0(\underline{r})$.

[a] This discussion is partly based on a previous article published in Lecture Manuscripts of the 37^{th} Spring School of the Institute of Solid State Research [6].

[b] Compared to a function $f(x)$, which is defined as a mapping from a variable x to a number f, a functional $F[f(x)]$ is defined as a mapping from a function $f(x)$ to a number F.

If one assumes that two potentials v_{ext} and v'_{ext}, which differ by more than a constant, lead to the same ground-state wavefunction Ψ_0, the subtraction of the Schrödinger equations for v_{ext} and v'_{ext} gives

$$(v_{ext} - v'_{ext})|\Psi_0\rangle = (E - E')|\Psi_0\rangle \; . \tag{2}$$

In regions with $\Psi_0 \neq 0$ the constant value $E - E'$ implies that the two potentials v_{ext} and v'_{ext} can differ only by a constant. Thus the assumption that v_{ext} and v'_{ext} differ by more than a constant can only be satisfied in regions where Ψ_0 vanishes. However, regions (with nonzero measure), where Ψ_0 vanishes, cannot exist because the unique continuation theorem[7,8] states that Ψ_0 vanishes everywhere if Ψ_0 vanishes in a region of nonzero measure. Thus the assumption that two potentials v_{ext} and v'_{ext}, which differ by more than a constant, lead to the same ground-state wavefunction requires that this wavefunction vanishes everywhere which is clearly impossible.

If one assumes that two different (apart from a trivial phase factor) ground-state wavefunctions Ψ_0 and Ψ'_0 for the different potentials v_{ext} and v'_{ext} lead to the same ground-state density $n_0(\underline{r})$, one obtains (see appendix)

$$\langle \Psi'_0 | v_{ext} - v'_{ext} | \Psi'_0 \rangle = \int n_0(\underline{r})[v_{ext}(\underline{r}) - v'_{ext}(\underline{r})]\mathrm{d}\underline{r} \tag{3}$$

and

$$\langle \Psi_0 | v'_{ext} - v_{ext} | \Psi_0 \rangle = \int n_0(\underline{r})[v'_{ext}(\underline{r}) - v_{ext}(\underline{r})]\mathrm{d}\underline{r} \; . \tag{4}$$

From $\langle \Psi_0 | \hat{H}_{v'} | \Psi_0 \rangle > \langle \Psi'_0 | \hat{H}_{v'} | \Psi'_0 \rangle = E'_0$, where the strict larger sign arises because Ψ'_0 is the ground-state wavefunction for the Hamiltonian $\hat{H}_{v'}$ which leads to the the ground-state energy ${E_0}'$, whereas Ψ_0, which differs from Ψ'_0 by more than a trivial phase factor leads to a larger energy, and from $\langle \Psi'_0 | \hat{H}_v | \Psi'_0 \rangle > \langle \Psi_0 | \hat{H}_v | \Psi_0 \rangle = E_0$, where the strict larger sign arises using similar arguments, one obtains

$$\langle \Psi_0 | \hat{H}_{v'} | \Psi_0 \rangle + \langle {\Psi_0}' | \hat{H}_v | {\Psi_0}' \rangle > E'_0 + E_0 \; . \tag{5}$$

Here the substitution $\hat{H}_{v'} = \hat{H}_v + v'_{ext} - v_{ext}$ in the first term and $\hat{H}_v = \hat{H}_{v'} + v_{ext} - v'_{ext}$ in the second term and the use of $\langle \Psi_0 | \hat{H}_v | \Psi_0 \rangle = E_0$ and $\langle {\Psi_0}' | \hat{H}_{v'} | {\Psi_0}' \rangle = E'_0$ leads to

$$E_0 + \langle \Psi_0 | v'_{ext} - v_{ext} | \Psi_0 \rangle + {E_0}' + \langle {\Psi_0}' | v_{ext} - v'_{ext} | {\Psi_0}' \rangle > {E_0}' + E_0 \tag{6}$$

By inserting (3) and (4), which are valid because of the assumption that the two different ground-state wavefunctions Ψ_0 and Ψ'_0 lead to the same ground-state density $n_0(\underline{r})$, one obtains $E_0 + {E_0}' > {E_0}' + E_0$, which is clearly a contradiction, and the assumption cannot be true. Consequently, two external potentials $v_{ext} \neq v'_{ext} + const$ cannot lead to the same ground-state density. Therefore, the ground-state density uniquely determines the external potential up to a trivial constant and thus via the many-electron Schrödinger equation uniquely the many-electron wavefunctions of the system. This means that all stationary observables of the many-electron system are uniquely determined by the ground-state density. Unfortunately, for most physical properties it is not known how they can be calculated directly from the ground-state density without using the many-electron Schrödinger equation so that the unique determination is not often of practical use.

To calculate the ground-state energy E_0 the unique energy functional $E[n(\underline{r})]$ can be defined[10] by

$$E[n(\underline{r})] = \min_{\Psi \to n(\underline{r})} \langle \Psi | \hat{T} + \hat{U} + v_{ext} | \Psi \rangle = F[n(\underline{r})] + \int n(\underline{r}) v_{ext}(\underline{r}) \mathrm{d}\underline{r} \,, \tag{7}$$

where the minimum is over all wavefunctions, which give the density $n(\underline{r})$. The functional

$$F[n(\underline{r})] = \min_{\Psi \to n(\underline{r})} \langle \Psi | \hat{T} + \hat{U} | \Psi \rangle \tag{8}$$

does not depend on the external potential v_{ext} and but only on $\hat{T}$ and $\hat{U}$ and is universal in the sense that it is same for all systems described by the Schrödinger equation (1). From (7) one obtains the inequality

$$E[n(\underline{r})] \leq \langle \Psi | \hat{T} + \hat{U} + v_{ext} | \Psi \rangle \tag{9}$$

for all wavefunctions Ψ, which give the density $n(\underline{r})$. For the ground-state wavefunction Ψ_0 with the ground-state density $n_0(\underline{r})$ this means $E[n_0(\underline{r})] \leq \langle \Psi_0 | \hat{T} + \hat{U} + v_{ext} | \Psi_0 \rangle = E_0$. Since $\langle \Psi | \hat{T} + \hat{U} + v_{ext} | \Psi \rangle \geq E_0$ is valid for all wavefunctions because of the Rayleigh-Ritz minimum principle, this inequality is also valid for the wavefunction which leads to the minimum in (7). This means $E[n(\underline{r})] \geq E_0$ is valid for all densities, in particular for the ground-state density: $E[n_0(\underline{r})] \geq E_0$. Together with $E[n_0(\underline{r})] \leq E_0$ this shows $E_0 = E[n_0(\underline{r})]$ which proves the second part of the Hohenberg-Kohn theorem: the minimum of $E[n(\underline{r})]$ is obtained for the ground-state density and this minimum gives the ground-state energy

$$E_0 = \min_n E[n(\underline{r})] \,. \tag{10}$$

Here the minimization is over all densities which arise from antisymmetric wavefunctions for N electrons.

2.2 Kohn-Sham Equations

The theory discussed above has transformed the problem of finding the minimum of $\langle \Psi | \hat{H} | \Psi \rangle$ for many-electron trial wavefunctions Ψ into the seemingly much more simple problem of finding the minimum of $E[n(\underline{r})]$ for trial densities $n(\underline{r})$ which depend on only three space variables. However, since the explicit form of the functional $F[n(\underline{r})]$ is not known, the theory is rather abstract. Here, the idea of Kohn and Sham[9], the introduction of a fictitious auxiliary non-interacting electron system with the same ground-state density is of extraordinary importance. Because the Hohenberg-Kohn theorem is valid for all interaction strengths (that is for all values of e^2), it is also valid for the choice $e^2 = 0$ which according to (1) describes a non-interacting system with $U(\underline{r}, \underline{r}') = 0$. By the Hohenberg-Kohn theorem the ground-state density uniquely determines the external potential in the non-interacting system. This potential is usually called the effective potential $v_{eff}(\underline{r})$. For the non-interacting system the total energy functional (7) can be written as

$$E[n(\underline{r})] = T_s[n(\underline{r})] + \int n(\underline{r}) v_{eff}(\underline{r}) \mathrm{d}\underline{r} \tag{11}$$

because the functional $F[n(\underline{r})]$ (for $e^2 = 0$) reduces to the kinetic energy functional $T_s[n(\underline{r})]$ of non-interacting electrons. For the non-interacting system with potential

$v_{eff}(\underline{r})$ the ground-state density $n_0(\underline{r})$ and the ground-state kinetic energy $T_s[n_0(\underline{r})]$ can be calculated exactly by

$$n_0(\underline{r}) = \sum_i |\varphi_i(\underline{r})|^2 \quad \text{and} \quad T_s[n_0(\underline{r})] = \sum_i \int \varphi_i^\star(\underline{r})(-\frac{\hbar^2}{2m}\nabla_{\underline{r}}^2)\varphi_i(\underline{r})\mathrm{d}\underline{r}\,, \tag{12}$$

where $\varphi_i(\underline{r})$ are the Kohn-Sham wavefunctions (orbitals), which are obtained by solving a single-particle Schrödinger equation

$$\hat{H}_s\varphi_i(\underline{r}) = \left[-\frac{\hbar^2}{2m}\nabla_{\underline{r}}^2 + v_{eff}(\underline{r})\right]\varphi_i(\underline{r}) = \epsilon_i\varphi_i(\underline{r})\,. \tag{13}$$

The sums in (12) are over the N wavefunctions with lowest values of ϵ_i. To apply this scheme, a useful expression for the effective potential $v_{eff}(\underline{r})$ must be found. The important achievement of Kohn and Sham was the suggestion to separate the unknown functional $F[n(\underline{r})]$ in (7) into a sum of known terms and into an unknown, hopefully much smaller rest which must be approximated. The energy functional is written as

$$E[n(\underline{r})] = T_s[n(\underline{r})] + \int n(\underline{r})v_{ext}(\underline{r})\mathrm{d}\underline{r} + \frac{e^2}{2}\iint \frac{n(\underline{r})n(\underline{r}')}{|\underline{r}-\underline{r}'|}\mathrm{d}\underline{r}\mathrm{d}\underline{r}' + E_{xc}[n(\underline{r})]\,, \tag{14}$$

where the term which contains density products describes the classical electron-electron interaction (Hartree interaction) and the last term is the exchange-correlation energy functional defined as

$$E_{xc}[n(\underline{r})] = F[n(\underline{r})] - T_s[n(\underline{r})] - \frac{e^2}{2}\iint \frac{n(\underline{r})n(\underline{r}')}{|\underline{r}-\underline{r}'|}\mathrm{d}\underline{r}\mathrm{d}\underline{r}'\,. \tag{15}$$

For the ground-state density comparison of (11) and (14) shows that

$$\int n(\underline{r})v_{eff}(\underline{r})\mathrm{d}\underline{r} = \int n(\underline{r})v_{ext}(\underline{r})\mathrm{d}\underline{r} + \frac{e^2}{2}\iint \frac{n(\underline{r})n(\underline{r}')}{|\underline{r}-\underline{r}'|}\mathrm{d}\underline{r}\mathrm{d}\underline{r}' + E_{xc}[n(\underline{r})]\,. \tag{16}$$

is valid except for an unimportant trivial constant. The functional derivative of (16) with respect to $n(\underline{r})$ is given by

$$v_{eff}(\underline{r}) = v_{ext}(\underline{r}) + e^2\int \frac{n(\underline{r}')}{|\underline{r}-\underline{r}'|}\mathrm{d}\underline{r}' \; + v_{xc}[n(\underline{r})](\underline{r})\,, \tag{17}$$

where the exchange-correlation potential

$$v_{xc}[n(\underline{r})](\underline{r}) = \frac{\delta E_{xc}[n(\underline{r})]}{\delta n(\underline{r})} \tag{18}$$

is defined for every point $\underline{r}$ as a functional of the density. Equations (12) and (13) are technically single-particle equations with a local effective potential $v_{eff}(\underline{r})$. This local potential makes density functional calculations simpler than Hartree-Fock calculations where the potential is non-local acting as $\int V_{HF}(\underline{r},\underline{r}')\varphi_i(\underline{r}')\mathrm{d}\underline{r}'$.

The effective potential (17) depends on the density, which in turn depends on the effective potential according to (12) and (13). These equations must be solved self-consistently, which can be achieved by iteration: starting with a reasonable trial density the effective potential is calculated by (17). Then (12) and (13) are solved to determine a new density which is used again in (17). This process is repeated until input and output density of an iteration agree within the required accuracy. The straightforward iteration usually leads

to oscillations with increasing amplitude. The oscillation can be damped by input-output mixing or by more sophisticated schemes[11]. From the behaviour of the eigenvalues of the functional derivative $f(\underline{r}, \underline{r}') = \delta E[n(\underline{r})]/\delta n(\underline{r}')$ it can be concluded[12] that the mixing process always converges to a stable solution if small enough mixing parameters are used, but many iterations may be needed.

The single-particle states φ_i and the single-particle energies ϵ_i obtained by solving (13) are properties of the *non-interacting auxiliary* system. In the interacting system they have no physical meaning and their interpretation as measurable quantities is not justified, although this interpretation is often adequate. A particular problem connected with the energies ϵ_i is that the eigenvalue gap between unoccupied and unoccupied states can differ considerably from the fundamental physical gap Δ in insulators and semiconductors. This gap is defined as $\Delta = [E(N+1) - E(N)] - [E(N) - E(N-1)]$ as the difference of the energies required for adding and removing one electron. Here $E(N)$, $E(N+1)$ and $E(N-1)$ are the ground-state total energies of the system with N, $N+1$ and $N-1$ electrons.

2.3 Approximations for the Exchange-Correlation Energy Functional

In principle, density functional theory is exact, but since all complications of the many-particle problem are hidden in the functional $E_{xc}[n(\underline{r})]$, which is not known explicitly, the success of density functional calculations depends on whether reasonable approximations for this functional can be found. A rather simple and remarkably good approximation is the replacement of the exact functional E_{xc} by

$$E_{xc}^{LDA}[n(\underline{r})] = \int n(\underline{r})\, \epsilon_{xc}^{LDA}(n(\underline{r}))\, \mathrm{d}\underline{r}\,, \tag{19}$$

the so-called local density approximation (LDA), where $\epsilon_{xc}^{LDA}(n)$ is a function (not a functional) of the density. For a homogeneous interacting electron system with constant density, the local density approximation is exact and $\epsilon_{xc}^{LDA}(n)$ can be determined as function of n by quantum mechanical many-body calculations. The exchange part $\epsilon_x^{LDA}(n)$ of $\epsilon_{xc}^{LDA}(n)$ is simple and given by

$$\epsilon_x^{LDA}(n) = -\frac{3e^2}{4}\left(\frac{3}{\pi}\right)^{1/3} n^{1/3}\,, \tag{20}$$

whereas the correlation part $\epsilon_c^{LDA}(n)$ is more difficult to calculate. Accurate results for $\epsilon_c^{LDA}(n)$ have been obtained by the quantum Monte Carlo method[13] and reliable parametrizations[14,15] for these results are available.

For systems with more inhomogeneous densities, the integrand in (19) can be generalized by using the gradient $\nabla n(\underline{r})$ of the density, for instance in the form,

$$E_{xc}^{GGA}[n(\underline{r})] = \int f(n(\underline{r}), \nabla n(\underline{r}))\, \mathrm{d}\underline{r}\,. \tag{21}$$

While the input ϵ_{xc}^{LDA} in (19) is unique, the function f in (21) is not and different forms have been suggested incorporating a number of known properties of the exact functional, for instance scaling and limit behaviours, or empirical parameters. A well tested numerical approximation is the generalized gradient approximation (GGA)[16–18], which for instance,

improves the cohesive energies and lattice constants of the 3d transition metals. So-called meta-GGA functionals[19,20] were also proposed, where besides the local density and its gradient also other variables are introduced, for instance the kinetic energy density of the Kohn-Sham orbitals

$$E_{xc}^{meta-GGA}[n(\underline{r})] = \int f\left(n(\underline{r}), \nabla n(\underline{r}), \tau(\underline{r})\right) \mathrm{d}\underline{r} \quad \text{with} \quad \tau(\underline{r}) = \sum_i |\nabla \varphi_i(\underline{r})|^2 . \tag{22}$$

By the additional flexibility in (22) it has been possible to improve the accuracy compared to (21) for some physical properties without worsening the results for others.

Probably the most serious shortcoming of the exchange-correlation functionals presented above is that they do not provide a cancellation of the self-interaction arising from the classical Hartree term which is used in (14). This shortcoming is particularly problematic in systems with localized and strongly interacting electrons as transition metal oxides and rare earth elements and compounds. Several techniques have been suggested to deal with self-interaction problem. Perdew and Zunger[15] suggested to use a self-interaction corrected (SIC) functional, where the self-interaction is removed explicitly for each orbital. In the LDA+U method[21] explicit on-site Coulomb interaction terms are added. Another way to treat the problem is to use the so-called exact exchange expression

$$E_x^{KS}[n(\underline{r})] = -\sum_{ij} \iint \frac{\varphi_i^\star(\underline{r}')\varphi_i(\underline{r})\varphi_j^\star(\underline{r})\varphi_j(\underline{r}')}{|\underline{r} - \underline{r}'|} \mathrm{d}\underline{r}\mathrm{d}\underline{r}' \tag{23}$$

as part of energy functional. Note that $E_x^{KS}[n(\underline{r})]$ as well as $T_s[n(\underline{r})]$ given in (12) and $\tau(\underline{r})$ given in (22) are defined by the Kohn-Sham orbitals $\varphi_i(\underline{r})$. Nevertheless, they are still density functionals, since by (13) the orbitals are determined by the effective potential and thus by the density because of the Hohenberg-Kohn theorem. One problem[22] with the use of exact exchange is to treat correlation in a way which is compatible with the exchange (23). In chemistry hybrid functionals, for instance

$$E_{xc}^{hyb} = aE_x^{KS} + (1-a)E_x^{GGA} + E_c^{GGA} \tag{24}$$

as suggested by Becke[23,24], are rather popular, where the constant $a \approx 0.28$ is an empirical parameter. Another, even more popular example is the B3LYP (Becke[24], three-parameter, Lee-Yang-Parr[25]) exchange-correlation functional

$$E_{xc}^{B3LYP} = E_{xc}^{LDA} + a_0(E_x^{KS} - E_x^{LDA}) + a_x(E_x^{GGA} - E_x^{LDA}) + a_c(E_c^{GGA} - E_c^{LDA}) \tag{25}$$

which combines the exchange E_x^{KS} with exchange and correlation functionals of LDA and GGA type with three empirically fitted parameters. Technically, self-consistent calculations with E_x^{KS} are rather involved because the exchange potential v_x^{KS} defined as the functional derivative of $E_x^{KS}[n(\underline{r})]$ with respect to $n(\underline{r})$ is difficult to calculate[22].

2.4 Solution methods

Although in density functional theory only single-particle equations with a local potential must be solved, the required computations can be a challenging task, in particular for complex and large systems. Thus it cannot be considered as a surprise that the Nobel Prize in Chemistry 1998 was not only awarded to Walter Kohn “for his development of

the density functional theory", but also to John A. Pople "for his development of computational methods in quantum chemistry". Standard solution methods for the Kohn-Sham equation (13) usually apply an expansion of the single-particle wavefunctions in a set of basis functions and use the Rayleigh-Ritz variational principle to determine the expansion coefficients.

Historically, solution methods can be classified into three categories using plane waves, localized atomic(-like) orbitals or the atomic sphere concept. Plane waves are simple and a natural basis for periodic systems, but inadequate to represent the large variations of the low lying atomic core states so that plane waves usually require to replace the strong potential near the nuclei by a much weaker pseudopotential. Localized orbitals, for instance Gaussian, Slater or numerically constructed orbitals, are well suited to describe atomic-like features in molecules and solids and are widely used, in particular in chemistry. In atomic sphere methods different representations for the wavefunctions are used in the spheres around the atomic centers, where the wavefunctions rapidly vary particularly near the nuclei, and in the interstitial region between the spheres, where the wavefunctions behave smoothly. In the original atomic sphere methods, in Slater's augmented plane wave (APW) method and in the Korringa-Kohn-Rostoker (KKR) method this separation resulted in a complicated non-linear energy dependence. Here Andersen's development[26] of the linear augmented plane wave (LAPW) and the linear muffin-tin orbital (LMTO) method by linearizing the energy dependence was a real breakthrough for the use of atomic sphere methods.

A disadvantage of basis set methods is that, although the basis set (chosen by physical motivation) often yields acceptable results for a small number of basis functions, precise calculations can be rather costly because they may require a large number of basis functions. Due to these limitations, in recent years purely numerical methods have been developed to solve the Kohn-Sham (Schrödinger) equation, for instance by using finite differences,[27] finite elements,[28] multigrid[28–30] or wavelet[31,32] methods.

3 Linear Scaling

Although over the last decades the computational efficiency to solve the density functional equations has increased significantly, the system size which can be studied is still rather limited. Systems with a few hundred atoms can be treated routinely today, but larger systems with thousands of atoms require enormous computer resources, if standard techniques are used to solve the density functional equations. The main bottleneck is that the computing time in standard calculations increases with the third power of the number of atoms (electrons) in the system. Although the computing power has increased by a factor of ten every four years (Moore's law) in the past and one can expect a similar increase in the next years, one has to wait for more than a decade until a ten times larger system can be treated if standard density functional methods with their $O(N^3)$ behaviour of the computing time are used.

Since about ten years considerable effort has been spent to remove the $O(N^3)$ bottleneck in most or all parts of the computer codes for density functional calculations. Most of this work is based on a locality principle, the nearsightedness of electronic matter, which has been formulated in a series of papers by Kohn[33,34]. Another possibility is to exploit the inherent $O(N)$ capability of multigrid[35] and multiresolution[36] (wavelet) methods.

The nearsightedness principle means that in systems without long range electric fields (and for fixed chemical potential) the density change at a point r_0, which is caused by a potential change in a finite region far away (outside a sphere with radius R around r_0), is small and decays to zero if R increases to infinity. Thus the charge density in a region (for instance in the central region shown in Fig. 1) can be calculated from the potential in this region and from the potential in a surrounding buffer region, whereas the potential outside the buffer region can be neglected. This concept is directly exploited in divide and conquer techniques (see below).

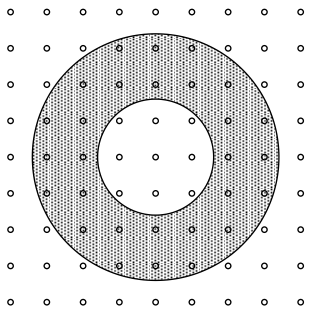


Figure 1. Schematic view of the central and the surrounding buffer region (in gray). The atomic positions are denoted by small circles.

A possibility to avoid the calculation of eigenstates which extend over the entire system is to work with the density matrix. For non-interacting particles the density matrix can be written in terms of the Kohn-Sham orbitals φ_i (the eigenstates of the non-interacting auxiliary system) for zero temperature as

$$\rho(\underline{r},\underline{r}') = \sum_i \varphi_i^\star(\underline{r})\varphi_i(\underline{r}') \tag{26}$$

and for non-zero temperature as

$$\rho(\underline{r},\underline{r}') = \sum_i f_i\varphi_i^\star(\underline{r})\varphi_i(\underline{r}') \;, \tag{27}$$

where the occupation numbers are given by $f_i = f((\epsilon_i - E_F)/kT)$. For $T = 0$ the sum is restricted to the occupied eigenstates, whereas for $T \neq 0$ all eigenstates are used. However, due to the decay of the Fermi-Dirac function $f(x) = (1 + \exp(x))^{-1}$ only low lying unoccupied states give appreciable contributions. In terms of the density matrix the density and kinetic energy given in (12) can be written as

$$n(\underline{r}) = \rho(\underline{r},\underline{r}) \quad \text{and} \quad T_s[n(\underline{r})] = \int \lim_{\underline{r}\to\underline{r}'} \left[-\frac{\hbar^2}{2m}\nabla_{\underline{r}}^2\rho(\underline{r},\underline{r}')\right] \mathrm{d}\underline{r}' \tag{28}$$

which shows that the effective potential (17) and all parts of the energy functional (14) can be calculated if $\rho(\underline{r},\underline{r}')$ is known. According to the nearsightedness principle the density matrix decays to zero for $|\underline{r} - \underline{r}'| \to \infty$. In insulators and semiconductors the decay is exponential for large distance[37–39]

$$\rho(\underline{r},\underline{r}') \sim \exp(-\gamma|\underline{r} - \underline{r}'|) \;, \tag{29}$$

whereas in metallic systems (at zero temperature) the decays is only algebraical

$$\rho(\underline{r},\underline{r}') \sim \frac{\cos(k_F|\underline{r}-\underline{r}'|)}{|\underline{r}-\underline{r}'|^2}\,, \tag{30}$$

where γ increases with the size of the band gap and $k_F = \sqrt{2mE_F}/\hbar$ denotes the Fermi wavevector.

3.1 Divide and Conquer Technique

A straightforward way to exploit the nearsightedness principle is to divide the system into overlapping subsystems and to solve the Kohn-Sham equations separately in each subsystem by standard methods taking into account an atom or a group of atoms in a central region surrounded by a buffer region. The density of the central regions is used and the density of the buffer regions is neglected. Examples of this approach are the divide and conquer technique proposed Yang[40] and the locally self-consistent multiple-scattering[41] (LSMS) or locally self-consistent Green function[42,43] (LSGF) methods which are based on KKR or LMTO calculations. Since each local interaction zone consisting of the central and its surrounding buffer region is treated independently, the effort in this approach scales linearly with system size and is parallelized easily over atoms or groups of atoms. A disadvantage is the limited accuracy[44] which can be achieved with a computationally affordable number of atoms in the local interaction zone since the effort increases cubically with this number.

3.2 Fermi Operator Expansion

The Kohn-Sham orbitals in (27) are eigenfunctions of the Hamilton operator $\hat{H}_s$ according to (13). From $\hat{H}_s\varphi_i = \epsilon_i\varphi_i$ one obtains $f((\hat{H}_s - E_F)/kT)\varphi_i = f_i\varphi_i$ and (27) can be written as

$$\rho(\underline{r},\underline{r}') = F(\hat{H}_s)\sum_i \varphi_i^\star(\underline{r})\varphi_i(\underline{r}') \tag{31}$$

with $F(\hat{H}_s) = f((\hat{H}_s - E_F)/kT)$. Since the sum in (31) is over all orbitals an arbitrary unitary transformation $\phi_i = \sum_j U_{ij}\varphi_j$ with $\sum_i U_{ik}^\star U_{ij} = \delta_{kj}$ can be used to rewrite (31) as

$$\rho(\underline{r},\underline{r}') = F(\hat{H}_s)\sum_i \phi_i^\star(\underline{r})\phi_i(\underline{r}') \tag{32}$$

This means that any complete set of basis functions can be used to evaluate the density matrix without the need to calculate explicitly the Kohn-Sham wavefunctions provided that one knows how to calculate $F(\hat{H}_s)\phi_i(\underline{r}')$. In the Fermi operator method[39,45] the Fermi function is expanded into Chebyshev polynomials so that $F(\hat{H}_s)$ is a polynomial in $\hat{H}_s$. Its action on the basis function $\phi_i(\underline{r}')$ is calculated according to the recursion relations of the Chebyshev polynomials by subsequent applications of $\hat{H}_s$. Linear scaling is obtained by neglecting the small elements of $F(\hat{H}_s)\phi_i(\underline{r}')$ which appear due to the exponential decay of the density matrix. Note that the use of Chebyshev polynomials requires that the eigenvalues of the Hamilton operator are in the interval $[-1,1]$ which can be achieved by shifting

and scaling. Similar in spirit to the Fermi operator expansion, which for $T \to 0$ corresponds to a polynomial expansion of a step function, is the kernel polynomial method[46] which uses a polynomial expansion of the δ function with factors designed to reduce the Gibbs oscillations arising from polynomial expansions of step or delta functions.

3.3 Recursion Method

The recursive application of a Hamilton operator to a basis set is also the essence of the recursion method[47,48] which is based on the Lanczos algorithm. The recursion method gives a continued fraction expansion for the density of states and for diagonal elements of the resolvent $E - \hat{H}_s$. It is used together with with divide and conquer approach, for instance, in the OpenMX program[49] based on a Krylov-subspace method[50,51].

3.4 Density Matrix Minimization

In the density matrix minimization approach[52–54] a direct minimization of the total energy with respect to the density matrix is performed. Here two constraints must be satisfied. The trial density matrix must give the correct number of electrons, $N = \int n(\underline{r})\mathrm{d}\underline{r} = \int \rho(\underline{r},\underline{r})\mathrm{d}\underline{r}$, and it must be idempotent $\hat{\rho}^2 = \hat{\rho}$ which means that

$$\int \rho(\underline{r},\underline{r}'')\rho(\underline{r}'',\underline{r}')\mathrm{d}\underline{r}'' = \rho(\underline{r},\underline{r}') \tag{33}$$

must be satisfied. This equation is equivalent to the requirement that all eigenvalues of the density matrix operator $\hat{\rho}$ are equal to one or zero. The constraint $N = \int \rho(\underline{r},\underline{r})\mathrm{d}\underline{r}$ can be treated by a Lagrange parameter which amounts to replacing the minimization of the total energy by minimization of the grand potential. The constraint of idempotency is taken into account by the "McWeeny purification"[55] which means to express $\hat{\rho}$ by $\hat{\rho} = 3\hat{\sigma}^2 - 2\hat{\sigma}^3$ with an auxiliary trial density matrix operator $\hat{\sigma}$. Provided that the trial operator $\hat{\sigma}$ has eigenvalues between -1/2 and 3/2, the eigenvalues of $\hat{\rho}$ are between 0 and 1 and the minimization process becomes a stable algorithm which drives the density matrix towards idempotency[52]. In the last years programs as CONQUEST[56] and ONETEP[57] have appeared which achieve linear-scaling by utilizing the decay of the density matrix[58–60].

3.5 Local Orbital Method

In the local orbital method[61–63] the Kohn-Sham energy functional is generalized by replacing (12) with

$$n(\underline{r}) = \sum_{ij} A_{ij}\phi_i^\star(\underline{r})\phi_j(\underline{r}) \quad \text{and} \quad T_s[n(\underline{r})] = \sum_{ij} A_{ij}\int \phi_i^\star(\underline{r})(-\frac{\hbar^2}{2m}\nabla_{\underline{r}}^2)\phi_j(\underline{r})\mathrm{d}\underline{r}\,, \tag{34}$$

where ϕ_i are non-orthogonal local orbitals. For $A_{ij} = \delta_{ij}$ this generalized functional agrees with the original Kohn-Sham functional and for $A_{ij} = S_{ij}^{-1}$, where $S_{ij} = \langle\phi_i|\phi_j\rangle$ is the overlap matrix, one obtains the correct functional for non-orthogonal orbitals. The problem with the choice $A = S^{-1}$ is that, whereas the overlap matrix is sparse for local

orbitals, its inverse is not sparse. To avoid the calculation of S^{-1} the local orbital method uses[61,63]

$$A = \sum_{k=0}^{n} (1 - S)^k \tag{35}$$

or the special choice[62] $n = 1$ which leads to $A = 2 - S$. During minimization the generalized functional approaches the correct one, but orthogonalization or calculation of the inverse of the overlap matrix, both requiring $O(N^3)$ operations, are avoided. Linear scaling with in the local orbital method is achieved by utilizing the decay of the density matrix[64], for instance within the SIESTA[65] program.

4 A Linear Scaling Algorithm for Metallic Systems

Since the density matrix decay in metals is only algebraical, an obvious idea is to make the decay faster by using a non-zero temperature. For $T \neq 0$ the density matrix in metals behaves for large distance as[36,37]

$$\rho(\underline{r}, \underline{r}', T) \sim \frac{\cos(k_F|\underline{r} - \underline{r}'|)}{|\underline{r} - \underline{r}'|^2} \exp(-\gamma|\underline{r} - \underline{r}'|) \,, \tag{36}$$

but it is not clear whether the decay constant γ, which is proportional to temperature, is large enough for reasonable temperatures so that linear scaling techniques developed for insulating systems can be applied also for metallic systems. Another difficulty for density matrix based techniques is that in metals no gap exists between occupied and unoccupied states so that an unambiguous choice of the states contributing to the density matrix is nontrivial. Nevertheless, some success has already been achieved for metallic systems[51,60].

Recently a linear scaling algorithm suitable for metals has been proposed in our institute[66,67]. This algorithm is based on the tight-binding (TB) version of the KKR Green function method[68,69] and on the electronic nearsightedness by exploiting a relation between finite-temperature density matrix and Green function. The principle of nearsightedness has been applied in KKR and LMTO calculations already for years, for instance for the embedding of impurities[70–72], where the fact is used that local potential perturbations lead to negligible density changes at large distance, and in the LSMS and LSGF methods discussed above. Compared to the LSMS and LSGF methods our algorithm seems to be more advantageous since in addition to the nearsightedness principle it also exploits the sparsity of the TB-KKR matrix. This sparsity alone leads already to an $O(N^2)$ behaviour of the computing time if the KKR matrix equations are solved by iteration.

4.1 Basic KKR Green Function Equations

Compared to wavefunction methods, where the density is calculated according to (12), the KKR Green function method obtains the density by

$$n(\underline{r}) = -\frac{2}{\pi} \mathrm{Im} \int_{-\infty}^{E_F} G(\underline{r}, \underline{r}; E) \mathrm{d}E \tag{37}$$

as an energy integral over the independent-particle Kohn-Sham Green function $G(\underline{r},\underline{r};E)$ which is defined as the solution of

$$\left[-\frac{\hbar^2}{2m}\nabla^2_{\underline{r}}+v_{eff}(\underline{r})-E\right]G(\underline{r},\underline{r}';E)=-\delta(\underline{r}-\underline{r}') \tag{38}$$

with the boundary condition $G(\underline{r},\underline{r}';E)\rightarrow 0$ for $|\underline{r}-\underline{r}'|\rightarrow\infty$. For the calculation of $G(\underline{r},\underline{r}';E)$ it is convenient to transform the differential equation (38) into an equivalent integral equation[69]

$$G(\underline{r},\underline{r}';E)=G^r(\underline{r},\underline{r}';E)+\int G^r(\underline{r},\underline{r}'';E)\left[v_{eff}(\underline{r}'')-v^r(\underline{r}'')\right]G(\underline{r}'',\underline{r}';E)\mathrm{d}\underline{r}'', \tag{39}$$

where v^r is the potential of a reference system, for which the Green function G^r is assumed to be known. This integral over all space is then divided into integrals over non-overlapping space-filling cells around the atomic positions $\underline{R}^n$. In each cell the multiple-scattering representation[69]

$$G(\underline{r}+\underline{R}^n,\underline{r}'+\underline{R}^{n'};E)=\delta^{nn'}G^n_s(\underline{r},\underline{r}';E)+\sum_{LL'}R^n_L(\underline{r};E)G^{nn'}_{LL'}(E)R^{n'}_{L'}(\underline{r}';E) \tag{40}$$

of the Green function is used, where L stands for the angular-momentum indices l and m and $\underline{r}$ and $\underline{r}'$ are cell-centred coordinates. With this representation the integral equation (39) can be solved by a matrix equation[69,73]

$$G^{nn'}_{LL'}(E)=G^{r,nn'}_{LL'}(E)+\sum_{n''L''L'''}G^{r,nn''}_{LL''}(E)\Delta t^{n''}_{L''L'''}(E)G^{n''n'}_{L'''L'}(E)\,. \tag{41}$$

Here the matrices have the dimension $N_{at}(l_{max}+1)^2$, where N_{at} is the number of atoms and l_{max} is the highest angular momentum l used (usually $l_{max}=3$ is sufficient). In (41) the Green function matrix elements $G^{nn'}_{LL'}(E)$ are the ones of the system and $G^{r,nn'}_{LL'}(E)$ are the ones of the reference system. These matrix elements are the only quantities in the KKR Green-function method which couple different atomic cells, whereas the single-scattering Green functions $G^n_s(\underline{r},\underline{r}';E)$ and wavefunctions $R^n_L(\underline{r};E)$ depend only on the potential $v_{eff}(\underline{r})$ inside cell n and the single-scattering t-matrix differences $\Delta t^n_{LL'}(E)$ only on the difference $v_{eff}(\underline{r})-v^r(\underline{r})$ of the potential and the reference potential inside cell n. All these single-scattering quantities can be calculated independently for each cell as described in[69,74] with a computational effort which naturally scales with the number of atoms. This means that for large systems the solution of (41) with its $O(N^3)$ computing effort requires by far the largest part of the computer resources, if the standard KKR Green function method is used, where due to free space as reference system the matrices in (41) are dense matrices.

Here the question is whether a reference system can be found, where the Green function matrix $G^{r,nn'}_{LL'}(E)$ is sparse, and whether the matrix equation (41) can be solved by iterative methods. This would reduce the computing effort from $O(N^3)$ to $O(N^2)$. Actually, only $O(N)$ elements of $G^{nn'}_{LL'}(E)$ with $n=n'$ are used for the density calculation, but in three-dimensional space the calculation of the $n=n'$ elements without the knowledge all other elements $G^{nn'}_{LL'}(E)$ seems to be impossible. In one-dimensional situations (e. g. for layered systems with two-dimensional periodicity) linear scaling algorithms to obtain the diagonal ($n=n'$) elements are known. Note that in one dimension the sparsity pattern of the Green

function matrix corresponds to a banded matrix with a bandwidth independent of the size of the system.

4.2 Repulsive Reference System

The standard reference system in the KKR method is free space. Here the Green function matrix elements $G_{LL'}^{0,nn'}(E)$ are analytically known, but decay rather slowly with distance between site n and n'. A reference system with exponentially decaying matrix elements can be obtained by using a repulsive potential. A useful reference system[68], where the matrix elements $G_{LL'}^{r,nn'}(E)$ can be calculated with moderate effort and without spoiling the rapid angular momentum convergence ($l \leq l_{max} = 3$), consists of an infinite array of repulsive potentials confined to nonoverlapping muffin-tin spheres around the sites $\underline{R}^n$ as it is schematically shown in Fig. 2. The matrix elements of this reference system, also called screened structure constants, can be calculated in real space by solving

$$G_{LL'}^{r,nn'}(E) = G_{LL'}^{0,nn'}(E) + \sum_{n''L''L'''} G_{LL''}^{0,nn''}(E) t_{L''L'''}^{r,n''}(E) G_{L'''L'}^{r,n''n'}(E) \qquad (42)$$

with reference t matrices $t_{L''L'''}^{r,n''}(E)$ which for each cell n'' are determined by the repulsive reference potential in this cell. Due to the rapid decay of $G_{L'''L'}^{r,n''n'}(E)$ with distance $|\underline{R}^{n''} - \underline{R}^{n'}|$, only a finite number N_{cl} of sites n'' contribute appreciably to the sum over n'' in (42). The neglect of more distant sites in (42) leads to a matrix equation of dimension $N_{cl}(l_{max}+1)^2$ which for each site n' can be solved independently. Setting exponentially small elements of $G_{LL''}^{r,nn''}(E)$ to zero makes this matrix sparse with a sparsity degree N_{cl}/N_{at} and reduces the computational effort to solve (41). The effort is then proportional to $N_{it}N_{cl}N_{at}^2$ instead of N_{at}^3 provided that (41) can be solved iteratively in N_{it} iterations.

4.3 Complex Energy Integration

One difficulty for the iterative solution of (41) is that iterations cannot converge at or near energies E, where the Green function $G(\underline{r}, \underline{r}'; E)$ has singularities. Such singularities appear on the real energy axis as poles (bound states) resembling the atomic core states and branch cuts (continuous eigenstates) resembling the valence and conduction bands. The difficulty is avoided if complex energies E with $\mathrm{Im}E \neq 0$ are used, which is straightforward in the KKR Green function method since the equations (38–42) are also valid for complex E. Moreover, since the Green function is an analytic function of E for $\mathrm{Im}E \neq 0$, the density (37) can be calculated by contour integration in the complex energy plane[75]. The necessarily real energy E_F at the end point of the contour is avoided by using the finite-temperature density functional formalism[76], where (37) is replaced by[69,77]

$$n(\underline{r}) = -\frac{2}{\pi}\mathrm{Im}\int_{-\infty}^{\infty} f(E - E_F, T) G(\underline{r}, \underline{r}; E)\mathrm{d}E\,. \qquad (43)$$

This integral can be calculated by a contour as shown in Fig. 2, where a typical set of integration mesh points is represented by crosses. The mesh points vertically above E_F correspond to singularities of the Fermi function (the so-called Matsubara energies) $E_j = E_F + (2j-1)\mathrm{i}\pi kT$ with $j = 1, 2, ...$ The other points are Gaussian integration points

constructed as described in Ref. 67. The contour starts on the negative real energy axis in the energy gap above the core and below the valence states. From there the contour goes parallel to the imaginary axis up to a chosen distance and then to infinity parallel to the real axis. The distance from the real axis is chosen as $2J\pi kT$, where J denotes the number of Matsubara energies at which the residues must be taken into account. Note that on the part of the contour, which is parallel to the real axis, the Fermi function is real as on the real axis due to its periodicity with period $2\mathrm{i}\pi kT$ and that practically no point with $\mathrm{Re}E > E_F$ exists because of the rapid decay of $f(E - E_F, T)$ for $\mathrm{Re}E > E_F$. The thick line in Fig. 2 along the real axis denotes the original integration of path of (37). The contour integration includes only contributions of valence states and the contributions of core states must be added separately.

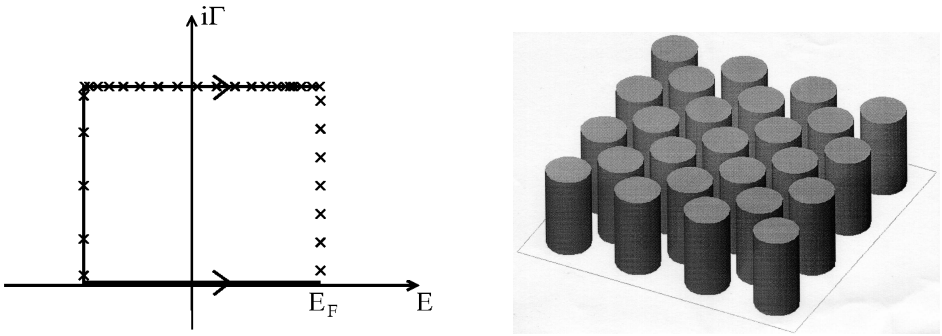


Figure 2. Integration contour in the complex energy plane with mesh points indicated by crosses (left picture) and a schematic view (in two dimensions) of a repulsive reference system with muffin-tin potentials of constant height (right picture).

4.4 Iterative Solution

Another difficulty for the iterative solution of (41) is that straightforward iteration, for instance in the form

$$G_{i+1}(E) = G^r(E) + G^r(E)\Delta t(E)G_i(E)\,, \tag{44}$$

which corresponds to the Born iterations in scattering theory, usually diverges. We found that convergent iterations can be obtained by Anderson mixing[11,78], which is used also sometimes to accelerate the density functional self-consistency. Anderson mixing combines input and output information of all previous iterations to prepare an optimal input for the next iteration. A disadvantage of Anderson mixing is that all this information must be kept which leads to large storage requirements. Alternatively, (41) can be solved iteratively by use of standard techniques which have been developed for systems of linear equations. With the TB-KKR matrix $M(E) = 1 - G^r(E)\Delta t(E)$, which for complex E is a complex non-Hermitian matrix, equation (41) can be written as a system of linear equations $M(E)G(E) = G^r(E)$. We found that the quasi-minimal-residual (QMR) method[79,80] in

its transpose free version is suitable to solve (41). The QMR method requires to store information from a few iterations and was better suited for the large supercells considered below than Anderson mixing.

An important feature of iterative solution is that each atom n' and each angular momentum component L' in (41) can be treated independently so that iterative solution is ideally suited for massively parallel computing. The independent treatment of each atom is in the spirit of the divide and conquer approach discussed above, however, whereas the divide and conquer approach usually implies an approximation, in our method the independent treatment is exact. For all systems studied so far, we could make the total-energy deviation compared to direct solution of (41) as small as we wanted, always smaller than 1 μeV using enough iterations.

4.5 Green Function Truncation

In order to arrive at an $O(N)$ algorithm the nearsightedness of electronic matter[33,34], which is the basis of most other linear-scaling methods, can be used in the following manner. From the relation

$$\rho(\underline{r},\underline{r}',T) = -\frac{2}{\pi}\mathrm{Im}\int_{-\infty}^{\infty} f(E-E_F,T)G(\underline{r},\underline{r}';E)\mathrm{d}E \tag{45}$$

between the finite-temperature density matrix $\rho(\underline{r},\underline{r}',T)$ and the Green function $G(\underline{r},\underline{r}';E)$ and from the property that the Green function decays faster for energies E with larger imaginary part, it can be concluded (via the complex energy contour integration discussed above) that the decay of $\rho(\underline{r},\underline{r}',T)$ is mainly determined by the decay of $G(\underline{r},\underline{r}';E_F+\mathrm{i}\pi kT)$ at the first Matsubara energy. Thus a neglect of the Green function for large distances $|\underline{r}-\underline{r}'|$ corresponds to a neglect of the finite-temperature density matrix for similar distances.

Since the single-scattering wavefunctions in (40) are only multiplicative factors, a truncation of the Green function directly corresponds to a neglect of Green function matrix elements $G_{LL'}^{nn'}$ beyond a chosen distance d_{cut}, which means that in (41) only $O(N_{tr}N_{at})$ elements $G_{LL'}^{nn'}$ are non-zero instead of $O(N_{at}^2)$. This reduces the computational effort by a factor N_{tr}/N_{at} if multiplication with zero elements is avoided by appropriate storage techniques. Here N_{tr} is the number of atoms which are included in the truncation region defined by $|\underline{R}^n - \underline{R}^{n'}| < d_{cut}$. The total effort is then proportional to $N_{it}N_{cl}N_{tr}N_{at}$. This increases linearly with N since the number of atoms N_{at} increases as the number of electrons N and since N_{cl} and N_{tr} are fixed numbers and since N_{it} approaches a constant value for large systems (see next section).

4.6 Iteration Behaviour and Total Energy Results

To illustrate how the calculated total energy is affected by the Green function truncation and how the number of iterations depends on the truncation region, results calculated with our algorithm for a large Ni supercell are shown in Fig. 3. The supercell was constructed by repeating a simple cubic unit cell with four atoms 32 times in all three space directions resulting in a supercell with $4\times 32^3 = 131072$ atoms. The lattice constant a was chosen as 11.276 nm, which is 32 times the experimental lattice constant of Ni. The repulsive muffin-tin potentials in the reference system had a height of 8 Ryd and cluster with $N_{cl} = 13$ atoms

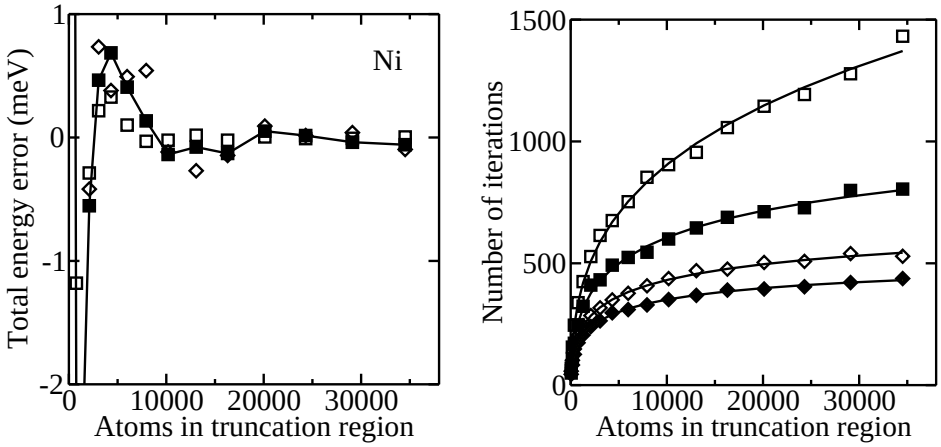


Figure 3. Left picture: Total energy error per atom as function of the number of atoms contained in the truncation region. Solid and open squares are for T = 800 K and 1600 K, diamonds for T = 400 K. The lines, which connect the results for T = 800 K, serve as guide for the eye. Right picture: Number N_{it} of iterations (matrix-vector multiplications averaged over the 16 angular momentum components) as function of the number N_{tr} of atoms contained in the truncation region. The lines are fitted to an exponential behaviour as described in the text. Solid (open) symbols denote results for the majority (minority) spin direction. The squares are for T = 800 K and the diamonds for T = 1600 K.

(central atom and its 12 nearest neighbours) were chosen to calculate the Green function matrix elements (42) of the reference system. A single point $\underline{k} = (1/4, 1/4, 1/4) \times 2\pi/a$ was used in the irreducible part of the Brillouin zone of the supercell. Since all atoms in the supercell are equivalent, the iterative solution of (41) was needed for only one value of n'. This represented an enormous reduction of the computational effort in the present model study compared to realistic systems with inequivalent atoms. Note that for $N_{at} = 131072$ and $l_{max} = 3$ the dimension of the matrices in (41) is $N_{at}(l_{max}+1)^2 = 2097152$ and that the matrix G^r has a sparsity degree of $13/131072 \approx 0.01\%$.

To study the truncation effect on the total energy one needs to know the total energy of Ni supercell calculated without truncation. Since only the density within one cell is required (all cells have the same density), such a calculation is possible with our present computer code. However, without truncation already about 7 Gigabyte are needed to store the non-zero elements of G^r and the self-consistent determination of the effective potential and the total energy would be rather expensive. Here the use of equivalent $\underline{k}$ point meshes[81] for the supercell and the simple cubic unit cell is of great help. If appropriate $\underline{k}$ points are used in the Brillouin zones, the calculated on-site Green function matrix elements for the supercell with equivalent atoms and for the simple cubic cell agree exactly. Thus the self-consistent potential and the correct total energy for the large Ni supercell with the single $\underline{k}$ point could be obtained inexpensively by self-consistent calculations for a simple cubic unit cell with 5984 $\underline{k}$ points.

The truncation regions were constructed by using more and more neighbour shells around the central atom so that always one more shell in the close-packed (110) direction was included. The smallest truncation region with 55 atoms included two neighbours in that direction and the largest truncation region with 34251 atoms included 18 neighbours in

that direction. The calculated total energy error is shown in Table 1 for small and in Fig. 3 for large truncation regions for three different temperatures. Whereas for small truncation regions the error can be as large as 0.1 eV, Fig. 3 shows that the error can be made smaller than 2 meV if truncation regions with a few thousand atoms are used. Since one is usually not interested in absolute total energies, but in total energy differences or derivatives (for instance forces, which can be calculated in the KKR method in a straightforward manner[69,82]), it can be expected that due to cancellation effects truncation regions with about a few hundred atoms are sufficient for the calculation of energy changes and forces with our linear scaling algorithm.

N_{tr}	ΔE_{400}	ΔE_{800}	ΔE_{1600}
55	121.9	135.6	123.2
177	87.3	105.8	125.6
381	33.6	31.0	26.4
767	-9.5	-7.9	-4.0
1289	-11.0	-9.8	-7.1
2093	-1.4	-1.9	-1.0

Table 1. Total energy error (in meV) as function of the number of atoms in the truncation region for three temperatures $T = 400$, 800, and 1600 K.

An important issue for our algorithm is how fast the iterations converge. The main computational work consists in matrix-vector multiplications, which were repeated independently for each angular-momentum component L' until the prescribed precision (specified by the residual norm $||r|| = 10^{-6}$ in the QMR method) were obtained. Fig. 3 shows the number of iterations (averaged over the $(l_{max}+1)^2 = 16$ angular-momentum components) at the first Matsubara energy $E_F + \mathrm{i}\pi kT$ where the slowest convergence exists. The number of iterations increases with increasing truncation region and can be fitted to an exponential behaviour of the form[66,67]

$$N_{it}(N_{tr}) = N_{it}^{\infty} - \alpha \exp(\gamma N_{tr}^{1/3}) \tag{46}$$

with three temperature dependent parameters N_{it}^{∞}, α and γ, which indicates that N_{it} approaches a constant value for large truncation regions. Whereas temperature has a pronounced effect on the computing time (via N_{it}), it seems that higher temperature does not much reduce the truncation error for the total energy, only for regions with more than 10000 atoms a reduction is seen. This probably means that the zero-temperature algebraical decay of the Green function (and density matrix) dominates the additional exponential decay caused by temperature up to truncation distances of approximately 10 times the Ni lattice constant.

Appendix

The expectation values $\langle\Psi|v_{ext}|\Psi\rangle$ and $\langle\Psi|\hat{U}|\Psi\rangle$ can be expressed in terms of the density $n(\underline{r})$ and the pair density $n_2(\underline{r},\underline{r}')$. The density is given by the expectation value of the

density operator $\hat{n}$ as

$$n(\underline{r}) = \langle\Psi|\hat{n}|\Psi\rangle = \int\cdots\int |\Psi(\underline{r}_1,...,\underline{r}_N)|^2 \sum_i^N \delta(\underline{r}-\underline{r}_i)\mathrm{d}\underline{r}_1...\mathrm{d}\underline{r}_N \,. \tag{47}$$

Multiplication with $v_{ext}(\underline{r})$ and integration leads to

$$\begin{aligned}\int n(\underline{r})v_{ext}(\underline{r})\mathrm{d}\underline{r} &= \int\cdots\int |\Psi(\underline{r}_1,...,\underline{r}_N)|^2 \sum_i^N \delta(\underline{r}-\underline{r}_i)v_{ext}(\underline{r}_i)\mathrm{d}\underline{r}_1...\mathrm{d}\underline{r}_N\mathrm{d}\underline{r} \\ &= \int\cdots\int |\Psi(\underline{r}_1,...,\underline{r}_N)|^2 \sum_i^N v_{ext}(\underline{r}_i)\mathrm{d}\underline{r}_1...\mathrm{d}\underline{r}_N \\ &= \langle\Psi|v_{ext}|\Psi\rangle \,.\end{aligned} \tag{48}$$

Here the first line arises by changing the argument in v_{ext} from $\underline{r}$ into $\underline{r}_i$, which is possible because of $\delta(\underline{r}-\underline{r}_i)$, and the second line arises by integration over the δ function. The pair density is given by the expectation value of the two-particle density operator $\hat{n}_2$ as

$$\begin{aligned}n_2(\underline{r},\underline{r}') = \langle\Psi|\hat{n}_2|\Psi\rangle = \int\cdots\int |\Psi(\underline{r}_1,...,\underline{r}_N)|^2 \sum_{i\neq j}^N\sum \delta(\underline{r}-\underline{r}_i)\delta(\underline{r}'-\underline{r}_j) \\ \mathrm{d}\underline{r}_1...\mathrm{d}\underline{r}_N \,.\end{aligned} \tag{49}$$

Proceeding similarly as above leads to

$$\begin{aligned}\int n_2(\underline{r},\underline{r}')U(\underline{r},\underline{r}')\mathrm{d}\underline{r}\mathrm{d}\underline{r}' &= \int\cdots\int |\Psi(\underline{r}_1,...,\underline{r}_N)|^2 \sum_{i\neq j}^N\sum \delta(\underline{r}-\underline{r}_i)\delta(\underline{r}'-\underline{r}_j) \\ &\quad \times U(\underline{r}_i,\underline{r}_j)\mathrm{d}\underline{r}_1...\mathrm{d}\underline{r}_N\mathrm{d}\underline{r}\mathrm{d}\underline{r}' \\ &= \int\cdots\int |\Psi(\underline{r}_1,...,\underline{r}_N)|^2 \sum_{i\neq j}^N\sum U(\underline{r}_i,\underline{r}_j)\mathrm{d}\underline{r}_1...\mathrm{d}\underline{r}_N \\ &= 2\int\cdots\int |\Psi(\underline{r}_1,...,\underline{r}_N)|^2 \sum_{i<j}^N\sum U(\underline{r}_i,\underline{r}_j)\mathrm{d}\underline{r}_1...\mathrm{d}\underline{r}_N \\ &= 2\langle\Psi|\hat{U}|\Psi\rangle \,,\end{aligned} \tag{50}$$

where the double sum over $i \neq j$ has been has replaced by twice the double sum over $i < j$. Note that the approximation $n_2(\underline{r},\underline{r}') = n(\underline{r})n(\underline{r}')$ leads to the expression of the electron-electron interaction used in (14) and the pair density must be distinguished one-particle density matrix defined as

$$\rho(\underline{r},\underline{r}') = N\int\cdots\int \Psi^\star(\underline{r},\underline{r}'_2,...,\underline{r}_N)\Psi(\underline{r}',\underline{r}'_2,...,\underline{r}_N)\mathrm{d}\underline{r}_2...\mathrm{d}\underline{r}_N \,. \tag{51}$$

References

1. L. H. Thomas, *The calculation of atomic fields*, Proc. Cambridge Philos. Soc. **23**, 542-548, 1927
2. E. Fermi, *Un metodo statistico per la determinazione di alcune proprietà dell'atomo*, Atti Accad, Naz. Lincei, Cl. Sci. Fis. Mat. Nat. Rend. **6**, 602–607, 1927.
3. P. Hohenberg and W. Kohn, *Inhomogeneous Electron Gas*, Phys. Rev. **136**, B864–B871, 1964.
4. R. Evans, Adv. Phys. **28**, 143–200 (1979) and in *Fundamentals of Inhomogeneous Fluids*, ed. by D. Hendersen (Dekker, New York, 1992).
5. M. Brack in *Density Functional Methods in Physics*, NATO ASI Series B, Vol. 123 (Plenum Press, New York, 1985).
6. R. Zeller, *Introduction to Density-Functional Theory*, in Computational Condensed Matter Physics, Lecture Manuscripts of the 37^{th} Spring School of the Institute of Solid State Research, S. Blügel, G. Gompper, E. Koch, H. Müller-Krumbhaar, R. Spatschek, R .G. Winkler (Eds.), Forschungszentrum Jülich GmbH, A1.1-A1.19, 2006.
7. E. H. Lieb, *Density Functionals for Coulomb Systems*, Int. J. Quant. Chem. **24**, 243–277, 1983
8. M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, (Academic, New York, 1978),Vol. **4**.
9. W. Kohn and L. J. Sham, *Self-Consistent Equations Including Exchange and Correlation Effects*, Phys. Rev. **140**, A1133–A1138, 1965.
10. M. Levy, *Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem*, Proc. Natl. Acad. Sci. U. S. A. **76**, 6062–6065, 1979.
11. V. Eyert, *A Comparative Study on Methods for Convergence Acceleration of Iterative Vector Sequences*, J. Comput. Phys. **124**, 271-285, 1996.
12. P. H. Dederichs and R. Zeller, *Self-consistency iterations in electronic-structure calculations*, Phys. Rev. B **28**, 5462, 1983.
13. D. M. Ceperly and B. J. Alder, *Ground State of the Electron Gas by a Stochastic Method*, Phys. Rev. Lett. **45**, 566–569, 1980.
14. S. H. Vosko, L. Wilk, and M. Nusair, *Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis*, Can. J. Phys. **58**, 1200–1211, 1980.
15. J. P. Perdew and A. Zunger, *Self-interaction correction to density functional approximations for many-electron systems*, Phys. Rev. B **23**, 5048–5079, 1981.
16. A. D. Becke, *Density-functional exchange-energy approximation with correct asymptotic behavior*, Phys. Rev. A **38**, 3098-3100, 1988.
17. J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, *Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation*, Phys. Rev. B **46**, 6671–6687, 1992; *Erratum*, Phys. Rev. B **48**, 4978, 1993.
18. J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*, Phys. Rev. Lett. **77**, 3865–3868, 1996.

19. J. P. Perdew, S. Kurth, A. Zupan, and P. Blaha, *Accurate Density Functional with Correct Formal Properties: A Step Beyond the Generalized Gradient Approximation*, Phys. Rev. Lett. **82**, 2544–2547, 1999.
20. J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, *Climbing the Density Functional Ladder: Nonempirical MetaGeneralized Gradient Approximation Designed for Molecules and Solids*, Phys. Rev. Lett. **91**, 146401-1–4, 2003.
21. V. I. Anisimov, J. Zaanen, and O. K. Andersen, *Band theory and Mott insulators: Hubbard U instead of Stoner I*, Phys. Rev. B **44**, 943–954, 1991.
22. S. Kümmel and L. Kronik, *Orbital-dependent density functionals: theory and applications*, Rev. Mod. Phys **80**, 3–60, 2008.
23. A. D. Becke, *A new mixing of Hartree-Fock and local density-functional theories*, J. Chem. Phys. **98**, 1372–1377, 1993.
24. A. D. Becke, *Density-functional thermochemistry. III. The role of exact exchange*, J. Chem. Phys. **98**, 5648–5652, 1993.
25. C. Lee, W. Yang, and R. G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density*, Phys. Rev. B **37**, 785–789, 1988.
26. O. K. Andersen, *Linear methods in band theory*, Phys. Rev. B **12**, 3060–3083, 1975.
27. J. R. Chelikowsky, N. Troullier, and Y. Saad, *Finite-difference-pseudopotential method: Electronic structure calculations without a basis*, Phys. Rev. Lett. **72**, 1240–1243, 1994.
28. S. R. White, J. W. Wilkins, and M. P. Teter, *Finite-element method for electronic structure*, Phys. Rev. B **39**, 5819–5833, 1989.
29. E. L. Briggs, D. J. Sullivan, and J. Bernholc, *Large-scale electronic-structure calculations with multigrid acceleration*, Phys. Rev. B **52**, R5471–R5474, 1995.
30. T. L. Beck, K. A. Iyer, and M. P. Merrick, *Multigrid methods in density functional theory*, Int. J. Quant. Chem. **341–348**, 1997, .
31. K. Cho, T. A. Arias, J. D. Joannopoulos, and P. K. Lam, *Wavelets in electronic structure calculations*, Phys. Rev. Lett. **71**, 1808–1811, 1993.
32. S. Wei and M. Y. Chou, *Wavelets in Self-Consistent Electronic Structure Calculations*, Phys. Rev. Lett. **76**, 2650–2653, 1996.
33. E. Prodan and W. Kohn, *Nearsightedness of electronic matter*, Proc. Natl. Acad. Sci. USA **102**, 11635–11638, 2005.
34. W. Kohn, *Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms*, Phys. Rev. Lett. **76**, 3168–3171, 1996.
35. T. L. Beck, *Real-space mesh techniques in density-functional theory*, Rev. Mod. Phys. **72**, 1041–1080, 2000.
36. T. A. Arias, *Multiresolution analysis of electronic structure: semicardinal and wavelet bases*, Rev. Mod. Phys. **71**, 267–311, 1999.
37. S. Goedecker, *Decay properties of the finite-temperature density matrix in metals*, Phys. Rev. B **58**, 3501–3502, 1998.
38. S. Ismail-Beigi and T. A. Arias, *Locality of the Density Matrix in Metals, Semiconductors, and Insulators*, Phys. Rev. Lett. **82**, 2127–2130, 1999.
39. S. Goedecker, *Linear scaling electronic structure methods*, Rev. Mod. Phys. **71**, 1085–1123, 1999.
40. W. Yang, *Direct calculation of electron density in density-functional theory*, Phys. Rev. Lett. **66**, 1438–1441, 1991.

41. Y. Wang, G. M. Stocks, W. A. Shelton, D. M. Nicholson, Z. Szotek, and W. M. Temmerman, *Order-N Multiple Scattering Approach to Electronic Structure Calculations*, Phys. Rev. Lett. **75**, 2867–2870, 1995.
42. I. A. Abrikosov, A. M. Niklasson, S. I. Simak, B. Johansson, A. V. Ruban, and H. L. Skriver, *Order-N Green's Function Technique for Local Environment Effects in Alloys*, Phys. Rev. Lett. **76**, 4203–4206, 1996.
43. I. A. Abrikosov, S. I. Simak, B. Johansson, A. V. Ruban, and H. L. Skriver, *Locally self-consistent Green's function approach to the electronic structure problem*, Phys. Rev. B **56**, 9319–9334, 1983.
44. A. V. Smirnov and D. D. Johnson, *Accuracy and limitations of localized Green's function methods for materials science applications*, Phys. Rev. B **64**, 235129-1–9, 2001.
45. S. Goedecker and L. Colombo, *Efficient Linear Scaling Algorithm for Tight-Binding Molecular Dynamics*, Phys. Rev. Lett. **73**, 122–125, 1993.
46. A. F. Voter, J. D. Kress, and R. N. Silver, *Linear-scaling tight binding from a truncated-moment approach*, Phys. Rev. B **53**, 12733–12741, 1996.
47. R. Haydock, V. Heine, and M. J. Kelly, *Electronic structure based on the local atomic environment for tight-binding bands*, J. Phys. C **5**, 2845–2858, 1972.
48. R. Haydock, V. Heine, and M. J. Kelly, *Electronic structure based on the local atomic environment for tight-binding bands. II*, J. Phys. C **8**, 2591–2605, 1975.
49. `http://www.openmx-square.org/`
50. T. Ozaki and H. Kino, *Efficient projector expansion for the ab initio LCAO method*, Phys. Rev. B **72**, 045121-1–8, 2005.
51. T. Ozaki, *O(N) Krylov-subspace method for large-scale electronic structure calculations*, Phys. Rev. B **74**, 245101-1–15, 2005.
52. X.-P. Li, R. W. Nunes, and D. Vanderbilt, *Density-matrix electronic-structure method with linear system-size scaling*, Phys. Rev. B **47**, 10891–10894, 1993.
53. M. S. Daw, *Model for energetics of solids based on the density matrix*, Phys. Rev. B **47**, 10895–10898, 1993.
54. E. Hernández and M. J. Gillan, *Self-consistent first-principles technique with linear scaling*, Phys. Rev. B **51**, 10157–10160, 1995.
55. R. Mc Weeny, *Some Recent Advances in Density Matrix Theory*, Rev. Mod. Phys. **32**, 335–369, 1960.
56. `http://www.conquest.ucl.ac.uk/index.html`
57. `http://www2.tcm.phy.cam.ac.uk/onetep/`
58. D. R. Bowler, T. Miyazaki, and M. J. Gillan, *Recent progress in linear scaling ab initio electronic structure techniques*, J. Phys.: Condens. Matter **14**, 2781–2798, 2002.
59. C-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, *Introducing ONETEP: Linear-scaling density functional simulations on parallel computers*, J. Chem. Phys. **122**, 084119-1–10, 2005.
60. C-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, *Using ONETEP for accurate and efficient $O(N)$ density functional calculations*, J. Phys.: Condens. Matter **17**, 5757–5769, 2005.
61. F. Mauri, G. Galli, and R. Car, *Orbital formulation for electronic-structure calculations with linear system-size scaling*, Phys. Rev. B **47**, 9973–9976, 1993.

62. P. Ordejón, D. A. Drabold, M. P. Grumbach, and R. M. Martin, *Unconstrained minimization approach for electronic computations that scales linearly with system size*, Phys. Rev. B **48**, 14646–14649, 1993.
63. J. Kim, F. Mauri, and G. Galli, *Total-energy global optimizations using nonorthogonal localized orbitals*, Phys. Rev. B **52**, 1640–1648, 1995.
64. J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejón, and D. Sanchez-Portal, *The SIESTA method for ab initio order-N materials simulation*, J. Phys.: Condens. Matter **14**, 2745–2779, 2002.
65. `http://www.uam.es/departamentos/ciencias/fismateriac/siesta/`
66. R. Zeller, *Towards a linear-scaling algorithm for electronic structure calculations with the tight-binding Korringa-Kohn-Rostoker Green function method*, J. Phys.: Condens. Matter **20**, 294215-1–8, 2008.
67. R. Zeller, *Linear-scaling total-energy calculations with the tight-binding Korringa-Kohn-Rostoker Green function method*, Phil. Mag. **88**, 2807–2815, 2008.
68. R. Zeller, P. H. Dederichs, B. Újfalussy, L. Szunyogh, and P. Weinberger, *Theory and convergence properties of the screened Korringa-Kohn-Rostoker method*, Phys. Rev. B **52**, 8807–8812, 1995.
69. N. Papanikolaou, R. Zeller, and P. H. Dederichs, *Conceptual improvements of the KKR method*, J. Phys.: Condensed Matter **16**, 2799–2823, 2002.
70. R. Zeller and P. H. Dederichs, *Electronic Structure of Impurities in Cu, Calculated Self-Consistently by Korringa-Kohn-Rostoker Green's-Function Method*, Phys. Rev. Lett. **42**, 1713–1716, 1979.
71. R. Podloucky, R. Zeller, and P. H. Dederichs, *Electronic structure of magnetic impurities calculated from first principles*, Phys. Rev. B **22**, 5777–5790, 1980.
72. O. Gunnarsson, O. Jepsen, and O. K. Andersen, *Self-consistent impurity calculations in the atomic-spheres approximation*, Phys. Rev. B **27**, 7144–7168, 1983.
73. R. Zeller, *Multiple-scattering solution of Schrödinger's equation for potentials of general shape*, J. Phys. C: Solid State Phys. **20**, 2347–2360, 1987.
74. B. Drittler, M. Weinert, R. Zeller, and P. H. Dederichs, *Vacancy formation energies of fcc transition metals calculated by a full potential Green's function method*, Solid State Commun. **79**, 31–35, 1991.
75. R. Zeller, J. Deutz, and P. H. Dederichs, *Application of the complex energy integration to selfconsistent electronic structure calculations*, Solid State Commun. **44**, 993–997, 1982.
76. N. D. Mermin, *Thermal Properties of the Inhomogeneous Electron Gas*, Phys. Rev. **137**, A1441–A1443, 1965.
77. K. Wildberger, P. Lang, R. Zeller, and P. H. Dederichs, *Fermi-Dirac distribution in ab initio Green's-function calculations*, Phys. Rev. B **52**, 11502–11508, 1995.
78. D. G. Anderson, *Iterative Procedures for Nonlinear Integral Equations*, J. Assoc. Comput. Mach. **12**, 547-560, 1965.
79. R. W. Freund and N. M. Nachtigal, *QMR: a quasi-minimal residual method for non-Hermitian linear systems*, Numer. Math **60**, 315–339, 1991.
80. R. W. Freund, *A Transpose-Free Quasi-Minimal Residual Algorithm for Non-Hermitian Linear Systems*, SIAM J. Sci. Comput. **14**, 470–482, 1993.
81. N. Chetty, M. Weinert, T. S. Rahman, and J. W. Davenport, *Vacancies and impurities in aluminum and magnesium*, Phys. Rev. B **52**, 6313–6326, 1995.

82. N. Papanikolaou, R. Zeller, P. H. Dederichs, and N. Stefanou, *Lattice distortion in Cu-based dilute alloys: A first-principles study by the KKR Green-function method*, Phys. Rev. B **55**, 4157–4167, 1997.

An Introduction to the Tight Binding Approximation – Implementation by Diagonalisation

Anthony T. Paxton

Atomistic Simulation Centre
School of Mathematics and Physics
Queen's University Belfast
Belfast BT1 7NN, UK
E-mail: Tony.Paxton@QUB.ac.uk

1 What is Tight Binding?

"Tight binding" has existed for many years as a convenient and transparent model for the description of electronic structure in molecules and solids. It often provides the basis for construction of many body theories such as the Hubbard model and the Anderson impurity model. Slater and Koster call it the tight binding or "Bloch" method and their historic paper provides the systematic procedure for formulating a tight binding model.[1] In their paper you will find the famous "Slater–Koster" table that is used to build a tight binding hamiltonian. This can also be found reproduced as table 20–1 in Harrison's book and this reference is probably the best starting point for learning the tight binding method.[2] Building a tight binding hamiltonian yourself, by hand, as in Harrison's sections 3–C and 19–C is certainly the surest way to learn and understand the method. The rewards are very great, as I shall attempt to persuade you now. More recent books are the ones by Sutton,[3] Pettifor[4] and Finnis.[5] In my development here I will most closely follow Finnis. This is because whereas in the earlier literature tight binding was regarded as a simple empirical scheme for the construction of hamiltonians by placing "atomic-like orbitals" at atomic sites and allowing electrons to hop between these through the mediation of "hopping integrals," it was later realised that the tight binding approximation may be directly deduced as a rigorous approximation to the density functional theory. This latter discovery has come about largely through the work of Sutton *et al.*[6] and Foulkes;[7] and it is this approach that is adopted in Finnis' book from the outset.

In the context of atomistic simulation, it can be helpful to distinguish schemes for the calculation of interatomic forces as "quantum mechanical," and "non quantum mechanical." In the former falls clearly the local density approximation (LDA) to density functional theory and nowadays it is indeed possible to make molecular dynamics calculations for small numbers of atoms and a few picoseconds of time using the LDA. At the other end of the scale, classical potentials may be used to simulate millions of atoms for some nanoseconds or more. I like to argue that tight binding is the simplest scheme that is genuinely quantum mechanical. Although you will read claims that the "embedded atom method" and other schemes are LDA-based, tight binding differs from these in that an explicit calculation of the electron *kinetic energy* is attempted either by diagonalising a hamiltonian, which is the subject of this lecture; or by finding its Green function matrix elements which is the subject of the lecture by Ralf Drautz.[8] The enormous advantage of the latter is that calculations scale in the computer linearly with the number of atoms, while diagonalisa-

tion is $\mathcal{O}(N^3)$. At all events, tight binding is really the cheapest and simplest model that can capture the subtleties in bonding that are consequences of the quantum mechanical nature of the chemical bond. Some well-known examples of these quantum mechanical features are magnetism, negative Cauchy pressures, charge transfer and ionic bonding; and of course bond breaking itself which is not allowed by simple molecular mechanics models. At the same time tight binding will reveal detailed insight into the nature of the bonds and origin of interatomic forces in the system you are studying.

1.1 The two centre approximation

In density functional calculations, the hamiltonian is constructed after making a choice of functions used to represent the wavefunctions, charge density and potential. If these are atom centred, for example gaussians, "fire balls" or Slater type orbitals rather than plane waves, then matrix elements of the hamiltonian may become spatial integrals of three such functions. An explicit formula taken from the LMTO method is displayed in equation (26) in section 3.2 below. This can be the most time consuming part of a bandstructure calculation, compared to the subsequent diagonalisation. In the tight binding approximation, we side step this procedure and construct the hamiltonian from a parameterised look up table. But the underlying theory has the same structure. Each hamiltonian matrix element is conceived as a integral of three functions, one potential and two orbitals centred at three sites. (We have made the *Ansatz* that the effective potential may be written as a sum of atom centred potentials.) If all are on the same site, this is a one centre, or *on-site* matrix element; if the orbitals are on different sites and are "neighbours" while the potential is on one of these sites we have a two centre matrix element, or "hopping integral." All other possibilities, namely three centre terms and overlap of orbitals on distant sites are neglected. This forms a central tenet of the tight binding approximation—the nearest neighbour, two centre approximation. The canonical band theory[9] allows us to isolate these terms explicitly and to predict under what circumstances these are indeed small (see section 3.2). The two centre approximation is more than just a convenient rejection of certain terms; it is implicit in the Slater–Koster table and in the calculation of interatomic force that the hamiltonian can be written in parameterised two centre form. This allows one to express the dependence of hopping integrals upon distance analytically. It is a feature of the quantum mechanical method that whereas the hamiltonian comprises short ranged two centre quantities only, the solution of the Schrödinger equation using this simple hamiltonian results in a density matrix that is possibly long ranged and includes many-atom interactions. Indeed the bond order potential exposes this many-atom expansion of the total energy explicitly.[8]

1.2 $\mathcal{O}(N^3)$ and $\mathcal{O}(N)$ implementations

The obvious way to tackle the tight binding electronic structure problem is the same as in density functional theory, namely by direct diagonalisation of the hamiltonian to obtain eigenvalues and eigenfunctions in the tight binding representation, section 2.1 below. This scales in the computer as the third power of the number of orbitals in the molecule or in the unit cell. In the solid state case one employs the Bloch theorem.[10] This means that one retains only the number of atoms in the primitive unit cell (rather than an infinite number) at the expense of having to diagonalise the hamiltonian at an infinite number of $\mathbf{k}$-points.

Luckily there is a well known and sophisticated number of ways to reduce this to a small number of points within the irreducible Brillouin zone.[11,12] The Bloch transform of a real space matrix $H_{\mathbf{R}L\,\mathbf{R}'L'}$ (in the notation described at equation (3) below) is

$$H_{\mathbf{R}L\,\mathbf{R}'L'}(\mathbf{k}) = \sum_{\mathbf{T}} H_{(\mathbf{R}+\mathbf{T})L\,\mathbf{R}'L'}\,\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{T}},$$

where $\mathbf{R}$ and $\mathbf{R}'$ run only over atoms in the primitive unit cell, while $\mathbf{T}$ are all the translation vectors of the lattice. As long as the matrix $H_{(\mathbf{R}+\mathbf{T})L\,\mathbf{R}'L'}$ is short ranged this can be done easily; for long ranged matrices such as the bare structure constants of (30) below, this must be done using the Ewald method. If you like you can *define* a two centre matrix as one for which the Bloch transformation can be reversed (using all $\mathcal{N}_{\mathbf{k}}$ points in the whole Brillouin zone)

$$H_{(\mathbf{R}+\mathbf{T})L\,\mathbf{R}'L'} = \frac{1}{\mathcal{N}_{\mathbf{k}}}\sum_{\mathbf{k}} H_{\mathbf{R}L\,\mathbf{R}'L'}(\mathbf{k})\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{T}}.$$

Indeed this is a way to extract a two centre tight binding hamiltonian from an LDA band-structure calculation;[13] an alternative approach is described in section 3.2 below.

In this lecture, I will concentrate solely on the method of direct diagonalisation, but an alternative and potentially much more powerful approach is to abandon $\mathbf{k}$-space, even for a periodic solid, and employ the recursion method to calculate not the eigenvalues and eigenfunctions of the hamiltonian H, but its greenian or Green function; formally for a complex variable z

$$\hat{G}(z) = (z - H)^{-1}.$$

Throwing away $\mathbf{k}$-space will lead to a huge computational benefit, namely that the calculation scales *linearly* with the number of orbitals, but there is a heavy price to pay—interatomic forces converge more slowly than the energy since they require off-diagonal greenian matrix elements and the sum rule derived in equation (16) below is not automatically guaranteed.[14,15] This can play havoc with a molecular dynamics simulation. The problem has been solved by the *bond order potential* which leads to a *convergent* expansion of the tight binding total energy in one-atom, two-atom, three-atom… terms—a many-atom expansion. This is the subject of the lecture by Ralf Drautz in this workshop.[8]

2 Traditional Non Self Consistent Tight Binding Theory

2.1 Density operator and density matrix

The traditional non self consistent tight binding theory, as described, say, by Harrison,[2] is explained here by following Horsfield *et al.*[16,17] We use H^0 to denote the hamiltonian to indicate that this is the non self consistent approximation to density functional theory as it appears in the Harris–Foulkes functional[5]—the first two lines in equation (37) below. (We follow the usual practice of suppressing the "hat" on the hamiltonian operator.) Hence, H^0 is the sum of non interacting kinetic energy and the effective potential generated by some *input*, superposition of atom centred, spherical charge densities.[5] The hamiltonian possesses a complete set of orthogonal eigenfunctions by virtue of the time independent Schrödinger equation,

$$H^0\psi_n = \varepsilon_n\psi_n,$$

which we will write using Dirac's bra-ket notation as

$$H^0 \left|n\right\rangle = \varepsilon_n \left|n\right\rangle . \tag{1}$$

ε_n are the eigenvalues of the hamiltonian and these are used to construct the *band energy*, $E_{\rm band}$, thus

$$E_{\rm band} = \sum_n f_n \, \varepsilon_n . \tag{2}$$

Here, f_n are *occupation numbers*. In an insulator or molecule assuming spin degeneracy these are either zero or two depending on whether ε_n is greater than or less than the Fermi energy. In a metal or molecule having a degenerate highest occupied level these are set equal to twice the Fermi function or some other smooth function having a similar shape.[12] As with any electronic structure scheme, if this is implemented as a *bandstructure* program and hence the hamiltonian is Bloch-transformed into $\mathbf{k}$-space, then the eigenstates are labelled by their band index and wave vector so that in what follows, the index n is to be replaced by a composite index $n\mathbf{k}$. (At the same time matrices become complex and you may assume that what follows until the end of this subsection applies separately at each $\mathbf{k}$-point.)

Central to the tight binding approximation is the expansion of the eigenstates of H^0 in a *linear combination of atomic(-like) orbitals* (LCAO). This means that we decorate each atomic site, which we denote $\mathbf{R}$ to label its position vector with respect to some origin, with orbitals having angular momentum $L = \ell m$. In this way, ℓ labels the orbitals as s, p or d character, while the L label runs as s, x, y, z, xy and so on. These orbitals may be written in bra-ket notation as

$$\left|\mathbf{R}L\right\rangle = \left|i\right\rangle \tag{3}$$

so that we can abbreviate the orbital site and quantum numbers into a single index i or j, k, l. In this way we have

$$\left|n\right\rangle = \sum_i c_i^n \left|i\right\rangle = c_i^n \left|i\right\rangle \tag{4}$$

and we use the famous Einstein summation convention, for brevity, whereby a summation over the indices i, j, k, l is understood if they appear repeated in a product. (Conversely we use n and m to label eigenstates of H^0 in equation (1) and these are not summed implicitly.) The expansion coefficients c_i^n are the eigenvectors of H^0 in the LCAO representation. The parameters of the tight binding model are the matrix elements of the hamiltonian in the LCAO basis which we write

$$H^0_{ij} = \left\langle i\right| H^0 \left|j\right\rangle .$$

We may *assume* that our chosen orbitals are orthogonal to each other, but to be more general there will a matrix of overlap integrals that may also comprise a part of our tight binding model. These are

$$S_{ij} = \left\langle i|j\right\rangle .$$

It then follows from (4) that (summing over j, remember)

$$\left\langle i|n\right\rangle = S_{ij} c_j^n \qquad \text{and} \qquad \left\langle n|i\right\rangle = \bar{c}_j^n S_{ji} \tag{5}$$

in which a "bar" indicates a complex conjugate. The Schrödinger equation (1) becomes a *linear eigenproblem*,

$$\left(H^0_{ij} - \varepsilon_n S_{ij}\right) c^n_i = 0. \tag{6}$$

In the case of an *orthogonal* tight binding model, we have $S_{ij} = \delta_{ij}$, otherwise we need to solve a generalised eigenproblem which is done by a Löwdin transformation. Denoting H^0_{ij} and S_{ij} in bold by matrices, we insert $\mathbf{S}^{-\frac{1}{2}}\mathbf{S}^{\frac{1}{2}}$ after the right parenthesis in (6) and multiply left and right by $\mathbf{S}^{-\frac{1}{2}}$:

$$0 = \left(\mathbf{S}^{-\frac{1}{2}}\mathbf{H}^0\mathbf{S}^{-\frac{1}{2}} - \varepsilon_n \mathbf{1}\right)\left(\mathbf{S}^{\frac{1}{2}}\mathbf{c}\mathbf{S}^{-\frac{1}{2}}\right) = \left(\tilde{\mathbf{H}} - \varepsilon_n \mathbf{1}\right)\mathbf{z},$$

which can be solved as an orthogonal eigenproblem, and recover $\mathbf{c}$ from $\mathbf{z}$ by back-substitution using the previously obtained Cholesky decomposition of $\mathbf{S}$. Now we have our eigenvectors c^n_i from which we construct a density matrix, which is central to the electronic structure problem. The density matrix provides us with the band energy, local "Mulliken" charges, bond charges (in the non orthogonal case)[5], bond orders,[4] interatomic forces, and in the case of time dependent tight binding the bond currents via its imaginary part.[18] The density operator $\hat{\rho}$ needs to have the following properties.

Property 1. Idempotency, meaning $\hat{\rho}^2 = \hat{\rho}$,

Property 2. Tr $\hat{\rho} = N$, the number of electrons,

Property 3. Tr $\hat{\rho}H^0 = \sum_n f_n\, \varepsilon_n = E_{\mathrm{band}}$, the band energy,

Property 4. Tr $\hat{\rho}\frac{\partial}{\partial\lambda}H^0 = \frac{\partial}{\partial\lambda}E_{\mathrm{band}}$, the Hellmann-Feynman theorem.

We know from quantum mechanics[19,20] that the one particle density operator is *defined* as

$$\hat{\rho} = \sum_n f_n\ |n\rangle\langle n|\,.$$

To find its representation in the LCAO basis, we first define a unit operator,

$$\hat{1} = |i\rangle\, S^{-1}_{ij}\, \langle j|\,. \tag{7}$$

To show that it *is* the unit operator, write

$$\begin{aligned}\langle n|n\rangle = 1 &= \langle n|i\rangle\, S^{-1}_{ij}\, \langle j|n\rangle \\ &= \bar{c}^n_k S_{ki} S^{-1}_{ij} S_{jl} c^n_l \\ &= \bar{c}^n_i S_{ij} c^n_j\end{aligned}$$

(after using (5) and swapping indices) which is consistent with (4). More generally we have

$$\langle n|m\rangle = \delta_{nm} = \bar{c}^n_i S_{ij} c^m_j. \tag{8}$$

Now using our unit vector, we write the density operator in our, possibly non orthogonal, LCAO basis,

$$\begin{aligned}\hat{\rho} = \sum_n f_n\ |n\rangle\langle n| &= \sum_n f_n\ \hat{1}\,|n\rangle\langle n|\,\hat{1} \\ &= \sum_n f_n\ |i\rangle\, c^n_i \bar{c}^n_j\, \langle j|\,.\end{aligned} \tag{9}$$

A matrix element of the density operator is

$$\begin{aligned}\rho_{kl} &= \sum_n f_n \ \langle k|i\rangle \, c_i^n \bar{c}_j^n \, \langle j|l\rangle \\ &= \sum_n f_n \, S_{ki} c_i^n \bar{c}_j^n S_{jl} \end{aligned} \tag{10}$$

and in an orthogonal basis this reduces to the familiar density matrix

$$\rho_{ij} = \sum_n f_n \, c_i^n \bar{c}_j^n .$$

If you are familiar with general relativity or non cubic crystallography then you may wish to view the matrix S_{ij} as the metric tensor that "raises" and "lowers" indices of covariant and contravariant vectors.[6,15,21] Finnis[5] makes this point by distinguishing between "expansion coefficients" and "matrix elements" of the density operator. In this way the expansion coefficients of the density operator in the LCAO basis are $\sum_n f_n \, c_i^n \bar{c}_j^n$, while to obtain density matrix elements their indices are "raised" by elements of the metric tensor as in (10); in the orthogonal case ($S_{ij} = \delta_{ij}$) this distinction vanishes.

Now we can demonstrate that $\hat{\rho}$ has the properties 1–4 above. The following is really included here for completeness as the student may not find it elsewhere in the literature. However, on a first reading you may skip to section 2.3 after looking at equations (11), (12), (13), (16) and (17).

Property 1. Idempotency follows immediately from (9).

Property 2. Tr $\hat{\rho} = N$. We must take the trace in the eigenstate basis, hence

$$\begin{aligned}\text{Tr}\,\hat{\rho} &= \sum_m \sum_n f_n \ \langle m|i\rangle \, c_i^n \bar{c}_j^n \, \langle j|m\rangle \\ &= \sum_m \sum_n f_n \, \bar{c}_k^m S_{ki} c_i^n \bar{c}_j^n S_{jl} c_l^m \\ &= \sum_m \sum_n f_n \, \delta_{mn}\delta_{nm} = \sum_n f_n \ = N. \end{aligned}$$

After the second line we have used (8). We can make partial, "Mulliken" charges q_i which amount to the occupancy of orbital i,

$$N = \sum_i q_i = \sum_n f_n \, \bar{c}_i^n S_{ij} c_j^n ,$$

using (8). Because of its importance in tight binding, we will write the Mulliken charge associated with orbital i explicitly,

$$q_i = \sum_n f_n \ \sum_j \bar{c}_i^n S_{ij} c_j^n \tag{11}$$

in which the sum over i implied by the summation convention is, in this instance, suppressed. This is a *weighted decomposition of the norm.* Note that in this and the following you can easily extract the simpler expressions for the more usual orthogonal

tight binding by replacing S_{ij} with the Kronecker δ_{ij} in the implicit sums, in which case

$$q_i = \sum_n f_n \ |c_i^n|^2 \, .$$

It is worthwhile to note that in an orthogonal tight binding model the total charge can be decomposed into individual atom centred contributions; on the other hand non orthogonality introduces *bond charge*[4] so that as seen in (11) there is a summation over both atom centred and bond charges. You may prefer the latter picture: we all know that in a density functional picture the covalent bond arises from the accumulation of charge in between the atoms; in an orthogonal tight binding model one might ask how is this accumulation described? The answer is that it is captured in the *bond order*.[4,8]

Property 3. $\mathrm{Tr}\ \hat{\rho}H^0 = \sum_n f_n \, \varepsilon_n = E_{\mathrm{band}}$.

$$\begin{aligned}
\mathrm{Tr}\ \hat{\rho}H^0 &= \sum_m \sum_n f_n \ \langle m|i\rangle \, c_i^n \bar{c}_j^n \, \langle j| \, H^0 \, |m\rangle \\
&= \sum_m \sum_n f_n \, \bar{c}_k^m S_{ki} c_i^n \bar{c}_n^j S_{jl} c_l^m \varepsilon_m \\
&= \sum_m \sum_n f_n \, \delta_{mn} \delta_{nm} \varepsilon_m = \sum_n f_n \, \varepsilon_n = E_{\mathrm{band}}
\end{aligned}$$

using (1). One may wish to construct partial band energies, E_i, in an equivalent way as

$$E_{\mathrm{band}} = \sum_i E_i = \sum_n f_n \, \bar{c}_i^n H_{ij} c_j^n .$$

The corresponding decomposition of the *bond energy* (18) in section 2.3 is the starting point of the many-atom expansion in the bond order potential.[8]

Property 4. The Hellmann–Feynman theorem tells us that

$$\frac{\partial}{\partial \lambda} \left(\mathrm{Tr}\ \hat{\rho}H^0 \right) = \mathrm{Tr}\ \hat{\rho} \frac{\partial}{\partial \lambda} H^0$$

because solution of the eigenproblem (6), through the Rayleigh–Ritz procedure leads us to a density matrix that is variational with respect to any parameter λ which may be, for example, a component of the position vector of an atom $\mathbf{R}$. Hence to calculate the interatomic force we need to find

$$\begin{aligned}
\mathrm{Tr}\ \hat{\rho} \frac{\partial}{\partial \lambda} H^0 &= \sum_m \sum_n f_n \ \langle m|i\rangle \, c_i^n \bar{c}_j^n \, \langle j| \, \frac{\partial}{\partial \lambda} H^0 \, |m\rangle \\
&= \sum_n f_n \, \bar{c}_i^n c_j^n \, \langle i| \, \frac{\partial}{\partial \lambda} H^0 \, |j\rangle \, .
\end{aligned}$$

Now our tight binding model furnishes us with hopping integrals, H_{ij}^0, and by employing a suitable scaling law, for example equation (23) below, the two centre approximation and the Slater–Koster table we will know how these depend on bond

lengths and angles; so while we don't actually know $\langle i| \frac{\partial}{\partial\lambda} H^0 |j\rangle$, the derivatives that we *do* know are

$$\frac{\partial}{\partial\lambda} H^0_{ij} = \frac{\partial}{\partial\lambda} \langle i| H^0 |j\rangle = \langle \frac{\partial}{\partial\lambda} i| H^0 |j\rangle + \langle i| H^0| \frac{\partial}{\partial\lambda} j\rangle + \langle i| \frac{\partial}{\partial\lambda} H^0 |j\rangle .$$

So

$$\mathrm{Tr}\, \hat{\rho} \frac{\partial}{\partial\lambda} H^0 = \sum_n f_n\, \bar{c}^n_i c^n_j \left[\frac{\partial}{\partial\lambda} H^0_{ij} - \langle \frac{\partial}{\partial\lambda} i| H^0 |j\rangle - \langle i| H^0| \frac{\partial}{\partial\lambda} j\rangle \right].$$

Now, to deal with the unknown last two terms, using (4)

$$\begin{aligned}
\sum_n f_n\, \bar{c}^n_i c^n_j \left[\langle \frac{\partial}{\partial\lambda} i| H^0 |j\rangle + \langle i| H^0| \frac{\partial}{\partial\lambda} j\rangle \right] &= \sum_n f_n \left[\bar{c}^n_i \langle \frac{\partial}{\partial\lambda} i|n\rangle \varepsilon_n + \varepsilon_n \langle n| \frac{\partial}{\partial\lambda} j\rangle c^n_j \right] \\
&= \sum_n f_n\, \varepsilon_n \left[\bar{c}^n_i c^n_j \langle \frac{\partial}{\partial\lambda} i|j\rangle + \bar{c}^n_i c^n_j \langle i| \frac{\partial}{\partial\lambda} j\rangle \right] \\
&= \sum_n f_n\, \varepsilon_n \bar{c}^n_i c^n_j \frac{\partial}{\partial\lambda} S_{ij}
\end{aligned}$$

since

$$\frac{\partial}{\partial\lambda} S_{ij} = \frac{\partial}{\partial\lambda} \langle i|j\rangle = \langle \frac{\partial}{\partial\lambda} i|j\rangle + \langle i| \frac{\partial}{\partial\lambda} j\rangle.$$

Finally we arrive at

$$\mathrm{Tr}\, \hat{\rho} \frac{\partial}{\partial\lambda} H^0 = \sum_n f_n\, \bar{c}^n_i c^n_j \left[\frac{\partial}{\partial\lambda} H^0_{ij} - \varepsilon_n \frac{\partial}{\partial\lambda} S_{ij} \right]. \tag{12}$$

2.2 Density of states and bond order

The *density of states* is central to electronic structure theory and is defined to be[22]

$$n(\varepsilon) = \sum_n \delta(\varepsilon - \varepsilon_n). \tag{13}$$

We can define a partial or *local* density of states, $n_i(\varepsilon)$, which is the density of states projected onto the orbital i. We write

$$\begin{aligned}
n(\varepsilon) &= \sum_n \langle n| \delta(\varepsilon - H^0) |n\rangle \\
&= \sum_n \sum_m \langle n|i\rangle\, S^{-1}_{ij}\, \langle j|m\rangle\, \langle m| \delta(\varepsilon - H^0) |n\rangle \\
&= \sum_n \sum_m \bar{c}^n_j S_{ji} S^{-1}_{ij} S_{jk} c^m_k\, \langle m| \delta(\varepsilon - H^0) |n\rangle \\
&= \sum_n \bar{c}^n_j S_{jk} c^n_k\, \langle n| \delta(\varepsilon - H^0) |n\rangle .
\end{aligned}$$

The first line follows from the Schrödinger equation (1) and in the second line we have inserted our unit operator (7) and a further unit operator, $\sum_m |m\rangle \langle m|$. The fourth line follows because of the orthogonality of the eigenvectors, $|n\rangle$ which means we have

$\langle m|\,\delta(\varepsilon - H^0)\,|n\rangle = \langle n|\,\delta(\varepsilon - H^0)\,|n\rangle\,\delta_{mn}$. Remember that in the fourth line j and k are dummy orbital indices to be summed over. We can replace these with i and j for neatness and this leads to

$$n(\varepsilon) = \sum_n \bar{c}_i^n S_{ij} c_j^n \delta(\varepsilon - \varepsilon_n) = \sum_i n_i(\varepsilon). \tag{14}$$

Writing the summation over j explicitly we see that the local density of states is

$$n_i(\varepsilon) = \sum_n \sum_j \bar{c}_i^n S_{ij} c_j^n \delta(\varepsilon - \varepsilon_n), \tag{15}$$

with no summation over i, and that this is a *weighted* density of states;[17] the weight in an orthogonal basis is simply $|c_i^n|^2$—compare this with the Mulliken decomposition (11).

An example is shown in figure 1. This is a self consistent *magnetic* tight binding calculation of the electronic structure of a Cr impurity in Fe, modelled as a dilute, ordered $Fe_{15}Cr$ alloy.[23] Very briefly magnetic tight binding is achieved by including a spin index, $|i\rangle = |\mathbf{R}L\sigma\rangle$, (now the occupation numbers vary between zero and *one*, not two) and adding an exchange potential to the self consistent hamiltonian to allow these to split. In addition to the Hubbard-U (see section 4) one includes a "Stoner I" parameter. We cannot go into details here, but it's gratifying that the simple tight binding model *quantitatively* reproduces the LSDA result, even to the extent of predicting the "virtual bound state" on the Cr impurity.[24,25]

The density of states can be used to find the band energy, since by the properties of the Dirac delta function,

$$\sum_n f_n \int \delta(\varepsilon - \varepsilon_n)\,\varepsilon\,\mathrm{d}\varepsilon = \sum_n f_n\,\varepsilon_n = E_{\mathrm{band}}.$$

If we allow the occupation numbers to be represented by the spin degenerate Fermi–Dirac distribution, $2f(\varepsilon)$, then we find, using (13) and our property 3, above,

$$E_{\mathrm{band}} = 2\int f(\varepsilon)\,\varepsilon\,n(\varepsilon)\,\mathrm{d}\varepsilon = \mathrm{Tr}\,\hat{\rho}H^0 \tag{16}$$

which is an important identity in tight binding theory and one which bears heavily on the convergence of the many atom expansion in the bond order potential.[26]

Finally in this section we should mention that the *bond order* which is central to the bond order potential[8] is obtained directly from the density matrix elements. We define

$$\Theta_{ij} = \frac{1}{2}\left(\rho_{ij} + \rho_{ji}\right)$$

as the *partial order of the bond* as contributed by orbitals i and j, it being understood that these are on different atomic sites. The bond order between sites $\mathbf{R}$ and $\mathbf{R}'$ is obtained by summing the partial bond order over all the orbitals on each atom in question,

$$\Theta_{\mathbf{R}\mathbf{R}'} = \sum_{LL'} \Theta_{\mathbf{R}L\,\mathbf{R}'L'}. \tag{17}$$

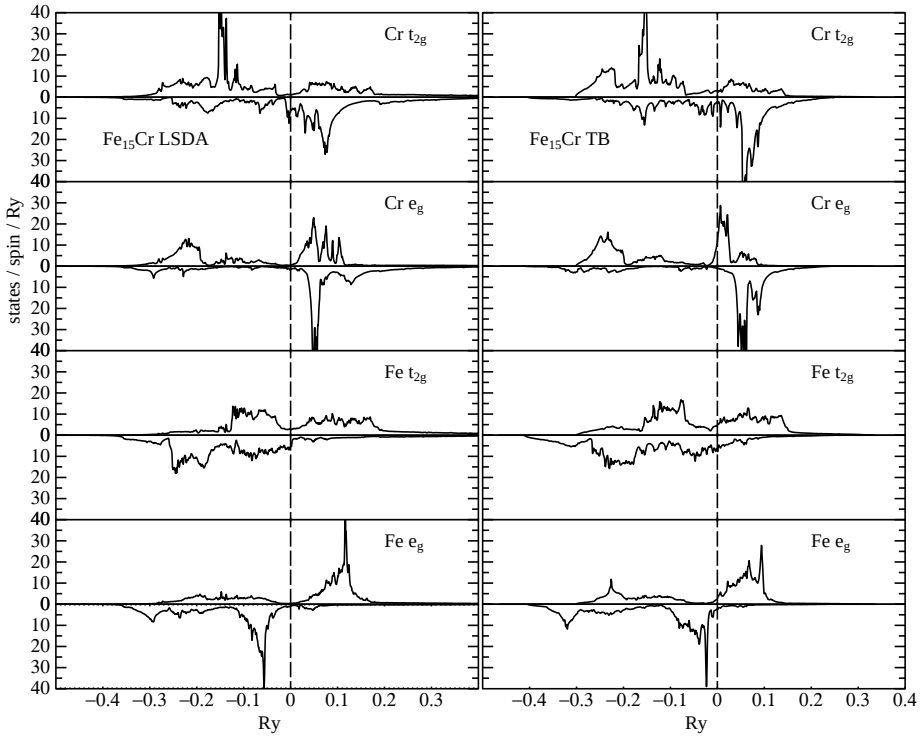


Figure 1. Example of a local density of states.[23] This is an ordered alloy, $Fe_{15}Cr$ on a body centred cubic (bcc) lattice. On the left is the local spin density functional result and on the right a simple, non orthogonal magnetic tight binding approximation. As is conventional, the spin up and down densities are shown as upright and upside down functions respectively. The Fe atom shown is the one closest to the Cr impurity and the density is projected onto the *d*-manifolds. Apart from the accurate description provided by the tight binding model, the most striking feature is the virtual bound state[24,25] seen as sharp peak in the local Cr density of states. It's notable that the *occupied*, spin up state has t_{2g} symmetry while its unoccupied partner belongs largely to the e_g manifold.

2.3 The tight binding bond model

Just as in density functional theory, the sum of occupied eigenvalues of the one electron hamiltonian is not the total energy. In the traditional tight binding approximation, beginning probably with the papers of Jim Chadi,[27] one writes simply

$$E_{\rm tot} = E_{\rm band} + E_{\rm pair}$$

for the total energy in the *band model* and $E_{\rm pair}$ is a pairwise repulsive energy whose functional form and parameters constitute ingredients of the tight binding model; it is intended to represent the double counting and ion–ion contributions to the density functional total energy.[27] "Double counting" is a term given to the electron–electron interaction energy in density functional theory. Because the theory is cast into a one electron form through the Kohn–Sham equations, the band energy, by summing over the eigenvalues, counts the electron–electron interaction twice. The interaction between, say, electrons in occupied states 1 and 2 is counted first when eigenvalue 1 is added in and again when eigenvalue 2 is

added. One cannot simply divide by two because $E_{\rm band}$ also contains kinetic and electron–ion energies which are not double counted. Hence one recalculates the electron–electron interaction energy and subtracts it, calling this the "double counting" correction.

Pursuing an argument that goes back as far as the sixties,[28,29] Pettifor[30] formulates the total energy in terms of the *bond* energy, $E_{\rm bond}$, rather than the band energy. The tight binding bond model[6] (TBBM) is the starting point for both self consistent tight binding which is described below in section 4 and for the modern bond order potentials.[8] Therefore we will pursue only the bond model further here. The essential point is that one arrives at the *covalent bond energy*[3,6] by removing the diagonal elements of $E_{\rm band} = {\rm Tr}\,\hat{\rho}H^0$. We recall that orbital indices i and j are a composite of site labels and quantum numbers, and write

$$E_{\rm bond} = \frac{1}{2}\sum_{\substack{ij\\ \mathbf{R}'\neq\mathbf{R}}} 2\rho_{ij}H^0_{ji} = \frac{1}{2}\sum_{\substack{\mathbf{R}L\,\mathbf{R}'L'\\ \mathbf{R}'\neq\mathbf{R}}} 2\,\rho_{\mathbf{R}L\,\mathbf{R}'L'}\,H^0_{\mathbf{R}'L'\,\mathbf{R}L}. \tag{18}$$

Here all terms are excluded from the double sum if orbitals i and j are on the same site $\mathbf{R}$. Note how by dividing and multiplying by two we can expose this as a sum of bond energies which is then divided by two to prevent each bond being double counted in the same way as a pair potential is usually written.

In the TBBM, the remaining diagonal terms in $E_{\rm band}$ are grouped with the corresponding quantities in the free atom. In the non self consistent tight binding approximation, the on-site matrix elements of H^0 are simply the free atom orbital energies (eigenvalues of the atomic hamiltonian)

$$H^0_{\mathbf{R}L\,\mathbf{R}L'} = \varepsilon_{\mathbf{R}\ell}\,\delta_{LL'}$$

and in addition to the hopping integrals, these are parameters of the tight binding model, ε_s, ε_p and ε_d. Furthermore, we assume certain orbital occupancies in the free atom, say, $N_{\mathbf{R}\ell}$, whereas after diagonalisation of the tight binding hamiltonian one finds these orbitals have occupancy given by the diagonal matrix elements of the density matrix. Hence there is a change in energy in going from the free atom limit to the condensed matter which is

$$\begin{aligned} E_{\rm prom} &= \sum_{\mathbf{R}L}\left(\rho_{\mathbf{R}L\,\mathbf{R}L}\,H^0_{\mathbf{R}L\,\mathbf{R}L} - N_{\mathbf{R}\ell}\,\varepsilon_{\mathbf{R}\ell}\right)\\ &= \sum_{\mathbf{R}L}\left(\rho_{\mathbf{R}L\,\mathbf{R}L} - N_{\mathbf{R}\ell}\right)\varepsilon_{\mathbf{R}\ell}\\ &= \sum_{\mathbf{R}L}\Delta q_{\mathbf{R}L}\,\varepsilon_{\mathbf{R}\ell}. \end{aligned} \tag{19}$$

We have assumed for now that on-site elements of H^0 are strictly diagonal and we recognise the first term in the first line as the difference between $E_{\rm band}$ and $E_{\rm bond}$. $E_{\rm prom}$ is called the *promotion energy* since it is the energy cost in promoting electrons that is very familiar, say, in the s–p promotion in covalent semiconductors in "preparing" the atoms in readiness to form the sp^3 hybrids in the diamond structure or the sp^2 hybrids in graphite. Thus in the tight binding bond model the *binding energy* is written as the total energy take away the energy of the free atoms,

$$E_{\rm B} = E_{\rm bond} + E_{\rm prom} + E_{\rm pair}. \tag{20}$$

The *interatomic force* is minus the gradient of the pairwise $E_{\rm pair}$ which is trivial, minus Tr $\rho\boldsymbol{\nabla}H^0$ which can be computed using equation (12) assuming that *on-site* hamiltonian matrix elements remain constant; this is the fully *non* self consistent tight binding approximation. And in fact at this level of approximation the band and bond models are indistinguishable. The first order variation of $E_{\rm B}$ with respect to atom cartesian coordinate R_α is

$$\frac{\partial}{\partial R_\alpha}E_{\rm B} = \sum_{\substack{\mathbf{R}L\,\mathbf{R}'L' \\ \mathbf{R}'\neq\mathbf{R}}} 2\rho_{\mathbf{R}L\,\mathbf{R}'L'}\,\frac{\partial}{\partial R_\alpha}H^0_{\mathbf{R}'L'\,\mathbf{R}L} + \frac{\partial}{\partial R_\alpha}E_{\rm prom} + \frac{\partial}{\partial R_\alpha}E_{\rm pair}. \qquad (21)$$

This is written for the orthogonal case, since this approximation forms a tenet of the TBBM. However, it's easy enough to add in the term from (12) containing the overlap and of course the diagonal elements S_{ii} are constant and do not contribute to the force. Note that the half in front of (18) has vanished—in the calculation of the force one sums over all bonds emanating from the atom at $\mathbf{R}$, not just half of them!

Now comes a rather subtle point. Unlike the band model, the bond model is properly consistent with the force theorem.[31] This states that there is no contribution to the force from self consistent redistribution of charge as a result of the virtual displacement of an atom. If a self consistent electronic system is perturbed to first order then that change in the bandstructure energy due to electron–electron interaction is exactly cancelled by the change in the double counting. This remarkable result means that by making a first order perturbation one cannot distinguish between an interacting and a non interacting electron system.[32] Indeed to calculate the interatomic force it is sufficient to find the change in band energy while making the perturbation—in this case the virtual displacement of an atom—in the frozen potential of the unperturbed system. In the band model there will be a first order change in the band energy upon moving an atom which *ought* to be cancelled by an appropriate change in the double counting, but *is not* because this is represented by the pair potential. Now we can discuss $\partial E_{\rm prom}/\partial R_\alpha$. In the band model there is no contribution to the force from $E_{\rm prom}$ (19); because of the variational principle $\varepsilon_{\mathbf{R}L}\delta q_{\mathbf{R}L} = 0$, and $q_{\mathbf{R}L}\delta\varepsilon_{\mathbf{R}L} = 0$ because the $\varepsilon_{\mathbf{R}L}$ are constants. However the Mulliken charge transfers are not necessarily zero and the force theorem does require any electrostatic contributions due to charge transfer to be included in the interatomic force;[33,34] neglect of these leads to the inconsistency of the band model. In the TBBM the most limited self consistency is imposed, namely the *Ansatz* of local charge neutrality so that electrostatic charge transfer terms vanish. This requires that for each site the *total* Mulliken charge difference between free atoms and condensed phase summed over all orbitals is zero. This is achieved iteratively by adjusting the on-site orbital energies. Here is the simplest example of a self consistent tight binding theory. It only affects the diagonal, on-site hamiltonian matrix elements and hence only $E_{\rm prom}$ is changed. Suppose we now write the hamiltonian as

$$H = H^0 + H' \qquad (22)$$

where H' has only diagonal elements which we may call $\Delta\varepsilon_{\mathbf{R}L}$. Then

$$\begin{aligned} E^{\rm TBBM}_{\rm prom} &= \sum_{\mathbf{R}L}\left(\rho_{\mathbf{R}L\,\mathbf{R}L} - N_{\mathbf{R}\ell}\right)H_{\mathbf{R}L\,\mathbf{R}L} \\ &= \sum_{\mathbf{R}L}\Delta q_{\mathbf{R}L}\left(\varepsilon_{\mathbf{R}\ell} + \Delta\varepsilon_{\mathbf{R}\ell}\right). \end{aligned}$$

In a sense this isn't really "promotion energy" anymore because we have applied the on-site energy shift to the free atoms also, but it is consistent with the formulation of the TBBM.[6] There will now be a contribution to the force on atom $\mathbf{R}$ from the new term $\sum_L \Delta q_L \Delta\varepsilon_\ell$. If the self consistency is achieved in such a way that all orbital energies are shifted by the same amount at each site, then this contribution vanishes because $\Delta\varepsilon_\ell$ is independent of L, moves to the front of the summation sign and $\sum_L \Delta q_L = 0$ by the local charge neutrality condition. Further and complete discussion of the TBBM can be found in the original paper[6] and in Finnis' book.[5]

3 How to Find Parameters

Now we turn to the question that is probably the most controversial. Many people dislike the tight binding approximation because whereas on the one hand we claim it to be close to the *ab initio* local density approximation solution, on the other we are reduced to finding parameters empirically just as if this were another classical potential. My own view is that *if* the tight binding approximation contains enough of the physics of the system we are studying then any reasonably chosen set of parameters will provide us with a useful model. From this point of view we would also demand that only a very small number of parameters is actually employed in the model. Furthermore it should be possible to choose these by intelligent guesswork and refinement starting from some well established set of rules; for example Harrison's solid state table,[2] or the prescription of Spanjaard and Desjonquères for the transition metals.[35] For example, the latter prescription has furnished us with useful tight binding models[23,36] for Mo, Re, Nb and Fe each with some five to ten adjustable parameters. Alternatively a 53-parameter model for Mo was produced by very careful fitting to a huge database of properties.[37] There doesn't appear to exist a particular advantage of one approach over the other and both types of model have turned out to be predictive of electronic and structural properties of the transition metals.

We need to distinguish between hamiltonian parameters—on-site orbital energies $\varepsilon_{\mathbf{R}\ell}$ and hopping integrals $H^0_{\mathbf{R}L\mathbf{R}'L'}$—and the parameters of the pair potential. Additional complications arise as described later in section 3.3 in the case of *environmentally dependent* parameters.[37]

I wish to illustrate the problem by reference to some examples from the literature.

3.1 Parameters by "adjustment"—example of ZrO_2

The tight binding model for zirconia,[38] ZrO_2, was designed to provide a description of the structural properties of this industrially important ceramic material. ZrO_2 suffers a number of structural phase transitions as a function of temperature. This is exploited in an extraordinary phenomenon called transformation toughening.[39] Its low temperature phase is monoclinic, at intermediate temperatures it is tetragonal and the high temperature modification is cubic. An open question was whether the tetragonal to cubic transition is of first or second order thermodynamically, order–disorder or displacive. Additionally, it is known that the cubic structure is stabilised at low temperature by doping with aliovalent cations (Y, Ca, Mg *etc*) while the mechanism for this was unknown. The tight binding model turned out to be capable of addressing both these issues and the order of the transition was discovered[40] as well as the mechanism of stabilisation of the cubic phase.[41] The

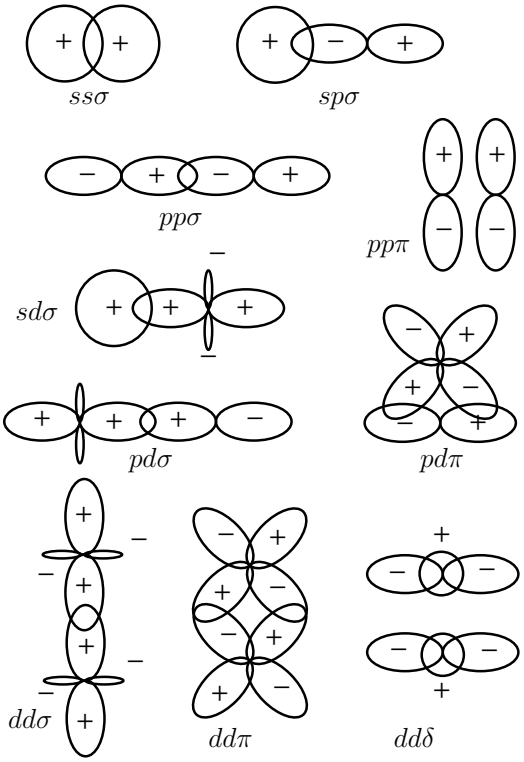


Figure 2. Bond integrals, after Majewski and Vogl.[42] This shows the well known atomic orbitals of the various s, p or d types joined along a bond. Radial symmetry along the bond is assumed leading to the designation of the bond as σ, π or δ. To construct a tight binding hamiltonian requires these fundamental bond integrals assembled through the Slater–Koster table using the direction cosines of the bond in a global cartesian system (these bond integrals are given with respect to a z-axis directed along the bond). This is illustrated in figure 6.5 in ref [3].

strategy of finding tight binding parameters was quite simple. Since the eigenvalues of the hamiltonian describe the energy bands it is sensible to adjust the on-site energies and hopping integrals to the LDA bandstructure, and then find a simple pair potential whose parameters are chosen to obtain, say, the equilibrium lattice constant and bulk modulus. In this case the smallest number of adjustable parameters was chosen to replicate the cubic phase in the hope that the model will then *predict* the ordering in energy of the competing phases. The steps are these.

1. Choose a *minimal* tight binding basis set. In this case d-orbitals were placed on the Zr atoms and s and p on the oxygen. We should mention that being an ionic crystal the TBBM is inadequate and this is in fact a *self consistent* tight binding model using polarisable ions. This is explained later in section 4. The hopping matrix elements are linear combinations of the fundamental bond integrals that are illustrated in figure 2. The particular linear combination depends on the bond angle geometry and is encapsulated in the Slater–Koster table.[1] This is illustrated in figure 6.5 in ref [3]. We only

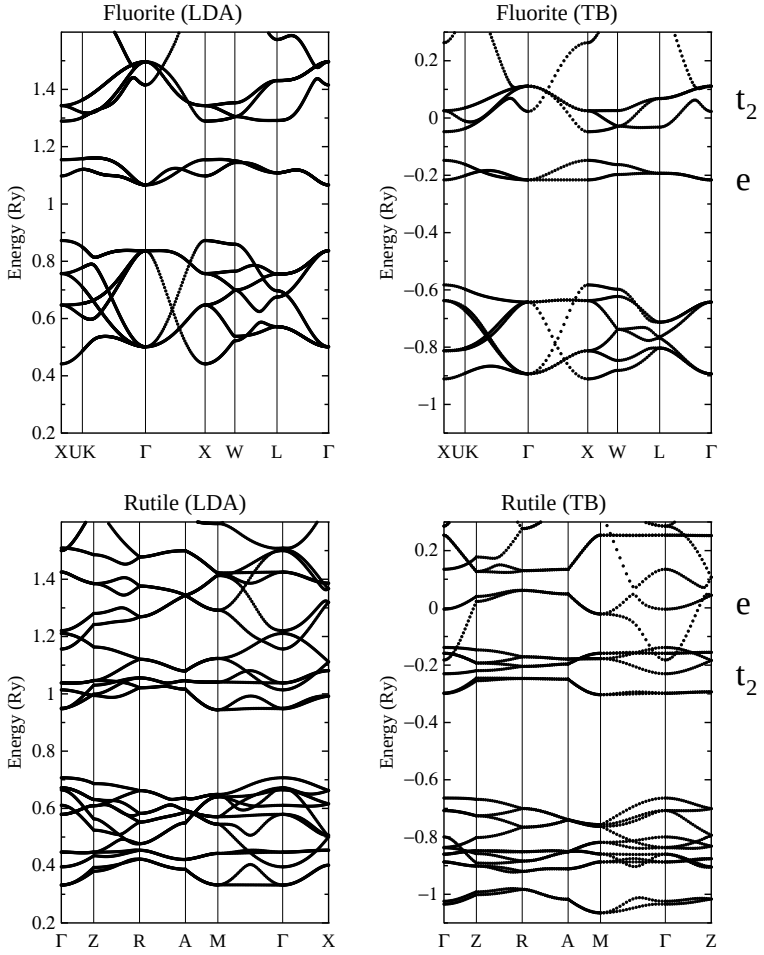


Figure 3. Energy bands of ZrO_2 using both LDA and a tight binding model in both fluorite and rutile crystal modifications. The model parameters were adjusted to the fluorite bands and the rutile bands are therefore a *prediction*. We should note that a number of features such as the splitting in the d-manifold into t_2 and e_g sub-bands and the crystal field widening of the p-derived ligand band in rutile are consequences of using the self consistent polarisable ion model, and this will be described later in section 4. But we can note in anticipation that it is the new Δ parameters that permit the ordering ($t_2 > e_g$) in the cubic crystal field and *vice versa* in the octahedral field to be reproduced automatically.

need to find the relevant fundamental bond integrals between neighbouring atoms. Zr–O first neighbour bonds require us to know $sd\sigma$, $pd\sigma$ and $pd\pi$ and we choose also to include second neighbour O–O bonds to be made by $ss\sigma$, $sp\sigma$, $pp\sigma$ and $pp\pi$ bond integrals. We have to choose both their value and the way in which they depend on bond length. There is a "canonical band theory," that is really appropriate for metals[9,43,44] but which *faux de mieux* we can apply more generally. This provides us with guidance on how the bond integrals decay with distance and also with certain ratios, namely $pp\sigma$:$pp\pi$ and $dd\sigma$:$dd\pi$:$dd\delta$, see equation (30) below. The required hopping

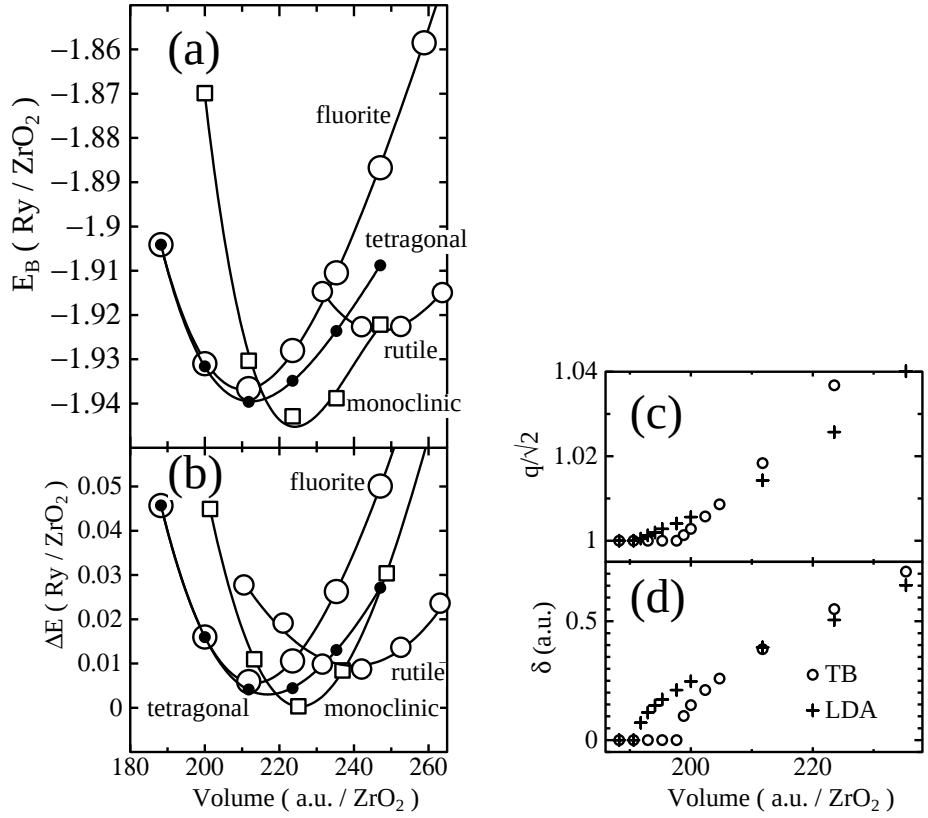


Figure 4. Total energy versus volume in four competing crystal structures of ZrO_2.[38] At each volume, the energy is minimised simultaneously with respect to all the remaining degrees of freedom. (a) LDA calculations of the absolute binding energy (energy with respect to spin polarised free atoms); (b) tight binding results referred to the equilibrium energy of the monoclinic phase. (c) and (d) show the axial ratio q and distortion parameter δ in the tetragonal modification as a function of volume.

integrals are initially taken from Harrison's solid state table and adjusted visually to obtain agreement with the shapes, and especially the widths of the LDA bands. One can also adjust to either the LDA or to experimental band gaps. Also the scaling of the bond integrals can be adjusted to the volume dependence of the LDA bandwidths.[a] The result is shown in figure 3.

We should give more detail of how the bond integrals depend on bond length, r. A very useful function is that of Goodwin, Skinner and Pettifor[46] (GSP)

$$(\ell\ell' m) = V_0 \left(\frac{d}{r}\right)^n \exp\left\{n\left[-\left(\frac{r}{r_c}\right)^{n_c} + \left(\frac{d}{r_c}\right)^{n_c}\right]\right\}. \qquad (23)$$

[a] It is very useful to have a computer program that can calculate energy bands, density of states, total energy using both LDA in some form and in the tight binding approximation, preferably all using the same input file. Luckily such a program exists.[45] Students may contact the author if they wish to learn how to use this.

Most important are the prefactor V_0 which is the value at the equilibrium bond length, d, and the exponent n which determines the slope of the function at equilibrium, since when $r = d$ the argument of the exponential vanishes. The role of n_c and r_c is to give a rapid decay to the function at around $r = r_c$.

2. A pair potential needs to be chosen. The GSP function can be used but in the ZrO_2 model a very simple Born–Mayer form was used between first neighbour Zr–O bonds only. The Born–Mayer function $\varphi(r) = A\,\mathrm{e}^{-br}$ has only two parameters which were fitted to the lattice constant and bulk modulus of cubic ZrO_2.

Figure 4 shows energy volume curves for the competing crystal structures comparing the tight binding model to LDA. Also shown are the order parameters that describe the tetragonal to cubic phase transition as functions of volume.

It is rather clear that the tight binding model for ZrO_2 gives a really excellent set of predictions, having been fitted (or adjusted, rather) only to the cubic structure. In particular the rutile structure is found to be much higher in energy than its competitors—a feature that cannot be reproduced in purely classical models. The vanishing of the order parameters with pressure is well reproduced qualitatively. This and the example shown in figure 1 where simple models, rather insensitive to the choice of parameters, reveal useful and predictive physics gives one confidence the tight binding approximation is indeed a valuable and reliable theory.

3.2 Parameters taken from first principles tight binding—example of Mo

Students who are not particularly interested in the details of an LMTO calculation, may skip this section after looking at figure 5 and subsequent comments. However section 3.3 *is* important. It makes sense to obtain the hamiltonian matrix elements from *ab initio* bandstructures. Probably the most transparent LDA bandstructure theory is the one provided by the linear muffin-tin orbitals (LMTO) method. In the atomic spheres approximation (ASA) the entire bandstructure problem is reduced to knowing four "potential parameters" in each $\mathbf{R}\ell$ site and angular momentum channel. Moreover these parameters have a clear interpretation in terms of the bandstructure. C is the centre of the band; Δ is the bandwidth parameter; γ is a distortion parameter describing the deviation from canonical bands and finally p is a small parameter allowing the eigenvalues to be correct up to third order in their deviation from some chosen energy called ε_ν. An LMTO is a composite orbital-like basis function. A sphere is inscribed about each atom with radius such that the sum of all sphere volumes equals the total volume; in a simple monatomic crystal this is the Wigner–Seitz radius. Within the sphere the radial Schrödinger equation is solved at the energy ε_ν in the current potential and this solution and its energy derivative are matched to solid Hankel and Bessel functions between the spheres. This matching condition is enough to provide the potential parameters which are functions of the logarithmic derivatives of the radial Schrödinger equation solutions $\phi_L(\mathbf{r}) = \phi_\ell(r)\,Y_L(\mathbf{r})$. Each LMTO envelope may be expanded about a given atomic site using the property that a Hankel function at one site may be written as a linear combination of Bessel functions at some other site. This property means that all the Hankel functions in the solid can be expressed as a "one centre" expansion about any one atomic sphere. The expansion coefficients are called "$\kappa = 0$ structure constants" and they transform under rotations according to the Slater–Koster table

and hence may be identified as $\ell\ell' m$ hopping integrals.[47,48] However conventional structure constants are very long ranged. To make contact with tight binding theory Andersen and Jepsen showed that one can make similarity transformations between sets of solid state LMTO's;[49] each basis set being equivalent to another since they give identical bandstructures. In particular Andersen demonstrated that one can defined a "most localised" and an "orthogonal" set of LMTOs. The transformation works like this. In the ASA an LMTO at site $\mathbf{R}$ is made up of a linear combination of a radial solution $\phi(\mathbf{r}-\mathbf{R})$ (the "head") and energy derivative functions $\dot{\phi}(\mathbf{r}-\mathbf{R}')$ ($\mathrm{d}\phi/\mathrm{d}\varepsilon$ evaluated at ε_ν) at all other sites (the "tails"). These are assembled into a one centre expansion using the structure constants. So an LMTO looks like this,

$$\chi_{\mathbf{R}L}(\mathbf{r}-\mathbf{R}) = \phi_{\mathbf{R}L}(\mathbf{r}-\mathbf{R}) + \sum_{\mathbf{R}'L'} \dot{\phi}_{\mathbf{R}'L'}(\mathbf{r}-\mathbf{R}')\, h_{\mathbf{R}'L'\,\mathbf{R}L}.$$

By a choice of normalisation, one can choose the $\dot{\phi}(\mathbf{r}-\mathbf{R}')$ to be those that are *orthogonal* to the radial solutions in each sphere. This particular set of energy derivative functions is given a superscript γ and one is said to be using the "γ-representation." More generally one can vary the normalisation by mixing in some radial solutions with the $\dot{\phi}(\mathbf{r}-\mathbf{R}')$ to make up the tails of the LMTO. To do this we write

$$\dot{\phi}_{\mathbf{R}L}(\mathbf{r}-\mathbf{R}) = \dot{\phi}^{\gamma}_{\mathbf{R}L}(\mathbf{r}-\mathbf{R}) + \phi_{\mathbf{R}L}(\mathbf{r}-\mathbf{R})\, o_{\mathbf{R}L}, \tag{24}$$

so that in the γ-representation, the potential parameter $o_{\mathbf{R}L}$ is zero. It's called o for overlap but has units of energy^{-1}. To construct the overlap matrix in the ASA one has to expand out $\langle\chi|\chi\rangle$; and similarly $\langle\chi\,|-\nabla^2+V_{\text{eff}}|\,\chi\rangle$ for the hamiltonian. If we write that $h_{\mathbf{R}'L'\,\mathbf{R}L}$ is an element of a matrix $\mathbf{h}$ and $o_{\mathbf{R}L}$ and $p_{\mathbf{R}L}$ are elements of diagonal potential parameter matrices, o and p, then Andersen finds for the overlap matrix[48]

$$\mathbf{S} = \mathbf{1} + o\mathbf{h} + \mathbf{h}o + \mathbf{h}p\mathbf{h}. \tag{25}$$

As we mentioned p is a small potential parameter. So in the γ-representation $o=0$ and to second order the overlap is unity and we have an orthogonal basis. The hamiltonian matrix turns out to be[48]

$$\mathbf{H} = \varepsilon_\nu + \mathbf{h} + \mathbf{h}o\varepsilon_\nu + \varepsilon_\nu o\mathbf{h} + \mathbf{h}\,(o+p\varepsilon_\nu)\,\mathbf{h}. \tag{26}$$

Again, in the γ-representation, neglecting third order terms the hamiltonian is just $\mathbf{H} = \varepsilon_\nu + \mathbf{h}$. So if one calculates structure constants and self consistent potential parameters using an LMTO code then one can build an orthogonal tight binding model by explicitly building $\mathbf{H}$ in the γ-representation. By construction, to second order it will reproduce the LDA energy bands.

Unfortunately there is no guarantee that this hamiltonian is short ranged. Andersen made a particular choice of the potential parameter $o_{\mathbf{R}L}$ by defining "screening constants" $\alpha_{\mathbf{R}L}$ in this way: ref [9], eq (91),

$$\frac{1}{o_{\mathbf{R}L}} = C_{\mathbf{R}L} - \varepsilon_{\nu,\mathbf{R}L} - \frac{\Delta_{\mathbf{R}L}}{\gamma_{\mathbf{R}L}-\alpha_{\mathbf{R}L}}. \tag{27}$$

They are called screening constants because the effect of adding radial solutions to the $\dot{\phi}^\gamma$ in (24) is to match the Schrödinger equation solutions in the sphere to Hankel functions $K_L(\mathbf{r}-\mathbf{R})$ that have been screened by additional Hankel functions at surrounding atomic

sites. There is an electrostatic analogy. The solid Hankel function represents the electrostatic potential due to a 2^ℓ multipole. This can be screened by surrounding the sphere with further grounded metal spheres, whose contribution to the potential is then provided by these further Hankel functions at the surrounding spheres. If one chooses the screening constants $\alpha_{\mathbf{R}L}$ to equal the band distortion parameters $\gamma_{\mathbf{R}L}$ then one arrives at the γ-representation since we get $o_{\mathbf{R}L} = 0$ in (27). All other representations are specified by choices of screening constants. The choice $\alpha_{\mathbf{R}L} = 0$ corresponds to the so called "first generation" LMTO which employs the standard $\kappa = 0$ KKR structure constants[b]

$$B_{\mathbf{R}'L'\,\mathbf{R}L} = -8\pi \sum_{L''} (-1)^\ell \frac{(2\ell''-1)!!}{(2\ell-1)!!(2\ell'-1)!!}\, C_{L'LL''}\, K_{L''}(\mathbf{R}-\mathbf{R}') \qquad (28)$$

where

$$K_L(\mathbf{r}) = r^{-\ell-1}\, Y_L(\mathbf{r}),$$

is the solid Hankel function,

$$C_{L''L'L} = \iint \mathrm{d}\Omega\, Y_{L''}\, Y_{L'}\, Y_L \qquad (29)$$

are Gaunt coefficients and Y_L are real spherical harmonics (see Appendix). The whole Slater–Koster table is encapsulated in this formula; the Gaunt coefficients provide selection rules that pick out certain powers of r and angular dependencies. By pointing a bond along the z-axis one can see how the canonical scaling and ratios come about since these structure constants are simply,[48]

$$\begin{aligned} B_{ss\sigma} &= -2/d \\ B_{sp\sigma} &= 2\sqrt{3}/d^2 \\ B_{pp\{\sigma,\pi\}} &= 6\{2,-1\}/d^3 \\ B_{sd\sigma} &= -2\sqrt{5}/d^3 \\ B_{pd\{\sigma,\pi\}} &= 6\sqrt{5}\{-\sqrt{3},1\}/d^4 \\ B_{dd\{\sigma,\pi,\delta\}} &= 10\{-6,4,-1\}/d^5 \end{aligned} \qquad (30)$$

in which d is a dimensionless bond length r/s, where s is conventionally chosen to be the Wigner–Seitz radius of the lattice. These can be compared with the cartoons in figure 2 in which the overlapping of two positive lobes leads to a negative bond integral and *vice versa*. This is because the orbitals are interacting with an attractive, negative, potential (section 1.1). Note how the factor $(-1)^\ell$ in (28) neatly takes care of the cases like $ps\sigma = -sp\sigma$. You have to be careful of these if you program the Slater–Koster table by hand.[5]

Transformations from the "first generation" to "second generation" LMTO basis sets are quite easily done. Having chosen screening constants one transforms the structure constants thus,[c]

$$\mathbf{B}^\alpha = \mathbf{B} + \mathbf{B}\alpha\mathbf{B}^\alpha \qquad (31)$$

[b] Andersen uses the symbol S for structure constants but we've already used it for the overlap, which is standard tight binding usage. Here we use B for Andersen's which differ by a prefactor $2/[(2\ell-1)!!(2\ell'-1)!!]$ and a minus sign from the KKR structure constants.[50]

[c] Clearly $B_{\mathbf{R}'L'\,\mathbf{R}L}$ has two centre form, section 1.1, as it depends only on the connecting vector $\mathbf{R}-\mathbf{R}'$ (28). It's less obvious that $\mathbf{B}^\alpha$ is a two centre matrix because of the three centre terms introduced by the second term

which is a Dyson equation, and α is a diagonal matrix. Then one transforms the potential parameters by defining a vector (we suppress the $\mathbf{R}L$ subscripts)

$$\xi = 1 + (C - \varepsilon_\nu)\,\frac{\alpha}{\Delta}$$

after which (ref [9], p. 88)

$$c = \varepsilon_\nu + (C - \varepsilon_\nu)\,\xi \;\; ; \qquad d = \xi^2 \Delta$$

where c and d are the band parameters C and Δ in the new representation. The overlap parameter o is transformed according to (27).

Andersen and Jepsen[49] determined empirically a set of screening constants, namely[d]

$$\alpha_s = 0.3485 \qquad \alpha_p = 0.05304 \qquad \alpha_d = 0.010714, \tag{32}$$

which lead to the "most localised" or most tight binding LMTO basis. Now one can construct hamiltonians and overlaps according to (26) and (25) by noting that the *first order* hamiltonian is constructed from potential parameters and structure constants[9,48]

$$h_{\mathbf{R}L\,\mathbf{R}'L'} = (c_{\mathbf{R}L} - \varepsilon_{\nu,\mathbf{R}L})\,\delta_{\mathbf{R}L\,\mathbf{R}'L'} + \sqrt{d_{\mathbf{R}L}}\,B^{\alpha}_{\mathbf{R}L\,\mathbf{R}'L'}\,\sqrt{d_{\mathbf{R}'L'}}.$$

Now we want our tight binding hamiltonian to have two centre form and it is easy to identify which are the three centre terms in the LMTO hamiltonian and overlap matrices—they are contained in the terms bilinear in $\mathbf{h}$, the last terms in (26) and (25). These terms (as do the linear terms) also contain two and one centre terms, of course, arising from the diagonal terms of $\mathbf{h}$. We can dispose of three centre terms in two ways.

1. We can work to *first order*, in which case, in both α- and γ-representations

$$\mathbf{H}^{(1)} = \varepsilon_\nu + \mathbf{h} \tag{33}$$

   and since $o\mathbf{h}$ terms are of second order, both these are orthogonal models with overlap being unity.

2. We can work to second order by retaining $o\mathbf{h}$ terms but neglecting the small potential parameter p^γ in the γ-representation. In this representation ($o = 0$) this is no different from the first order hamiltonian, and the overlap is unity. In the α-representation this introduces some additional two centre contributions to the matrix elements of the hamiltonian and overlap, and we are careful to extract one and two centre contributions from the last term in (26).

All this is illustrated in figure 5 for the bcc transition metal Mo. The screening constants from (32) are used. Here are some noteworthy points.

1. Clearly the two representations deliver different sets of hopping integrals. *You cannot expect density functional theory to furnish you with THE tight binding model.* On the other hand they show a proper decay with increasing bond length. The decay is

in (31). Nonetheless because the transformation is done in real space it is also a two centre matrix by virtue again of its dependence only upon $\mathbf{R} - \mathbf{R}'$. On the other hand it possesses additional "environmental dependence," see section 3.3.

[d] An alternative is to define $\alpha_{\mathbf{R}\ell} = (2\ell+1)(r_{\mathbf{R}\ell}/s)^{2\ell+1}$ by choosing site and ℓ-dependent "hard core radii" $r_{\mathbf{R}\ell}$.[51] This is consistent with "third generation LMTO."[52]

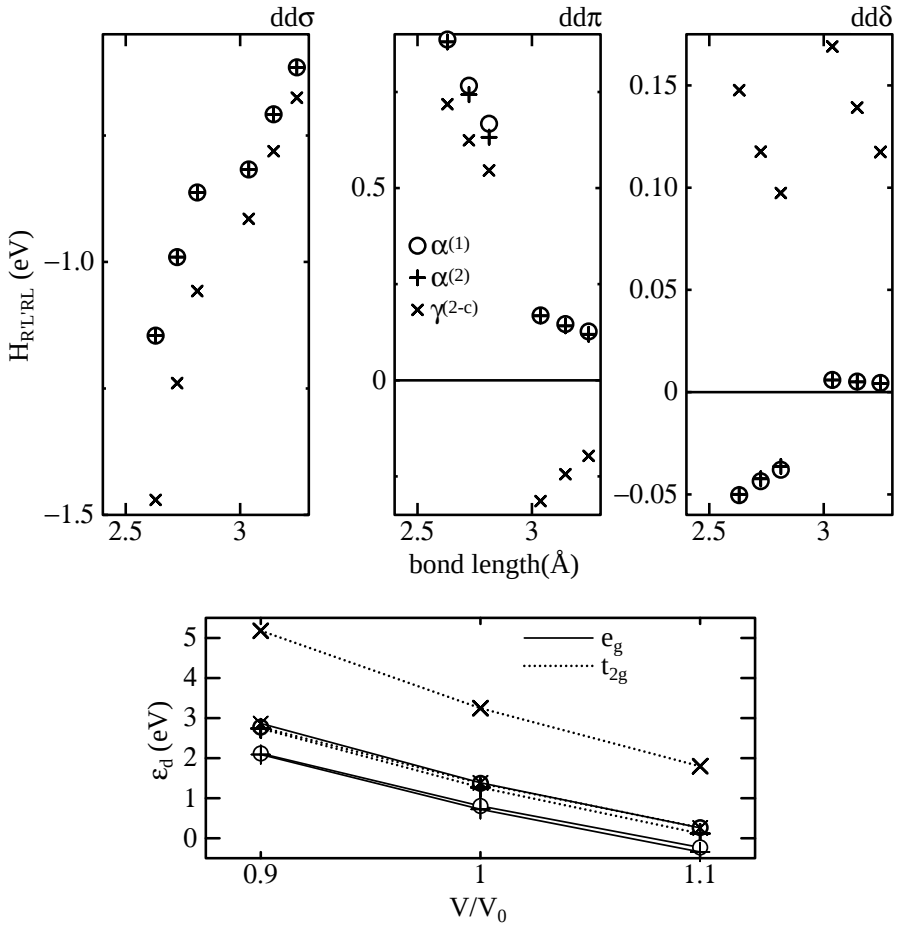


Figure 5. Hopping integrals $\ell\ell' m$ in the body centred cubic transition metal Mo, calculated using LMTO theory. The three integrals $dd\sigma$, $dd\pi$, and $dd\delta$ are found by rotating the z axis to first and then to second neighbour bonds and doing this at three different atomic volumes;[53] hence for each integral six values of $\ell\ell' m$ are shown as a function of bond length. Three model LMTO hamiltonians are used. The crosses refer to the two centre γ-representation; the circles to the *first order* α-representation and the pluses to the *second order*, two centre $\mathbf{H}^{\alpha}$. In the lower panel are shown the diagonal matrix elements and their, rather strong, volume dependence.

more rapid in the tight binding, α-representation as expected, furthermore the first order tight binding representation is strictly orthogonal; not shown in figure 5 are the overlap matrix elements in the second order tight binding representation, but indeed these are very small—no greater than 0.025 in magnitude. Note that the tight binding bond integrals respect the signs and roughly the canonical ratios of the bare structure constants (30) while in the γ-representation $dd\delta$ and the second neighbour $dd\pi$ have the "wrong" signs. Furthermore we would find that while the tight binding bond integrals shown reproduce the LDA bands using just first and second neighbour matrix elements, this is not the case for the γ-representation. Note that the first and second

order tight binding matrix elements are essentially the same; the additional second order terms may be safely neglected and the first order orthogonal hamiltonian (33) is clearly the proper one to use for this case.

2. If you have the patience, then you can do this exercise by hand in the case of the first order hamiltonian.[48] However the scheme has been automated and is implemented in our LMTO suite.[45]

3. Unfortunately the on-site energies shown in the lower panel of figure 5 are far from independent of volume. This is a remaining unsolved question for the construction of tight binding models in which the on-site energies are invariably constant (except of course for the adjustments in self consistent models to account for electrostatic shifts due to charge transfer, see (44) below). Andersen[48] points out that the Dyson equation (31) provides guidance on how to account for this volume dependence in terms of the local neighbour environment. Whereas on-site matrix elements of the bare structure constants, $\mathbf{B}$ are zero, we have from (31)

$$B^{\alpha}_{\mathbf{R}L\,\mathbf{R}L} = \sum_{\mathbf{R}''\neq\mathbf{R}} \sum_{L''} B_{\mathbf{R}L\,\mathbf{R}''L''}\; \alpha_{\mathbf{R}''\ell''}\; B^{\alpha}_{\mathbf{R}''L''\,\mathbf{R}L}$$

and the on-site matrix element of (33) is[51]

$$\varepsilon_{\mathbf{R}L} = c_{\mathbf{R}L} + d_{\mathbf{R}L} B^{\alpha}_{\mathbf{R}L\,\mathbf{R}L}.$$

However the band centre parameter c and bandwidth parameter d are also strongly volume dependent.[9,44] An important contrast with the ASA is that in tight binding, the on-site parameters are constant—the scaling law has to take care of both the bond length dependence at constant volume *and* the volume dependence itself.[54]

3.3 Environmentally dependent tight binding matrix elements

Possibly the most striking feature displayed in figure 5 is a discontinuity, most notably in the $dd\pi$ and $dd\delta$ bond integrals, between first and second neighbours. This is of particular importance to structures like bcc which have first and second neighbours rather similar in bond length. It means that one *cannot* find a simple scaling law, such as the GSP (23) that can connect all the points in the graph. This effect was noticed in the case of the $ss\sigma$ bond integral in Mo by Haas *et al.*[37] and they proposed a very significant development in tight binding theory, namely the use of *environmentally dependent* bond integrals.[55] The discontinuities in the dd bond integrals were noticed by Nguyen-Manh *et al.*[53] who offered the physical explanation in terms of "screening." The basic idea is that the bond between two atoms is *weakened* by the presence of a third atom. Therefore the scaling of a bond integral, say by the GSP function (23) is modified by multiplying it by $(1 - \mathcal{S}_{\ell\ell' m})$ where the "screening function," $\mathcal{S}_{\ell\ell' m}$, is the hyperbolic tangent of a function[37]

$$\xi^{\mathbf{R}\mathbf{R}'}_{\ell\ell' m} = A_{\ell\ell' m} \sum_{\substack{\mathbf{R}'' \\ \mathbf{R}''\neq\mathbf{R},\mathbf{R}'}} \exp\left[-\lambda_{\ell\ell' m}\left(\frac{|\mathbf{R}-\mathbf{R}''| + |\mathbf{R}'-\mathbf{R}''|}{|\mathbf{R}-\mathbf{R}'|}\right)^{\eta_{\ell\ell' m}}\right], \qquad (34)$$

in which A, λ and η are parameters to be fitted. This complicated expression can be simply explained.[37,53] As a third atom, $\mathbf{R}''$ approaches the $\mathbf{R} - \mathbf{R}'$ bond the term in parentheses becomes small, and approaches one in the limit that atom $\mathbf{R}''$ sits inside the $\mathbf{R} - \mathbf{R}'$

bond. This increases the value of the exponential and the tanh function smoothly reduces the $\mathbf{R} - \mathbf{R}'$ bond integral. Whereas Tang *et al.*[55] introduced this function empirically, Nguyen-Manh *et al.*[53] were able to derive its form using the theory of bond order potentials, and explain *why* $dd\sigma$ is not strongly screened while $dd\pi$ and $dd\delta$ are. Modern tight binding models[51,56,57] for transition metals are now fitted to curves such as those in figure 5 using (34). Indeed in these new schemes a repulsive energy is also fitted to an environmentally dependent function similar to (34). This is intended to make a better description of the valence–core overlap[44,58] between atoms which is short ranged but not pairwise and is otherwise not properly captured in the tight binding bond model. So nowadays one finds instead of (20)

$$E_{\mathrm{B}} = E_{\mathrm{bond}} + E_{\mathrm{prom}} + E_{\mathrm{env}} + E_{\mathrm{pair}} \tag{35}$$

in the TBBM, and E_{env} is the new environmentally dependent repulsive energy; it being understood that E_{bond} may be constructed using environmentally dependent hopping integrals too. E_{prom} is sometimes omitted,[56,57] in the instance that only one orbital angular momentum is included in the hamiltonian, for example if one employs a d-band model for transition metals.

4 Self Consistent Tight Binding

We described a tight binding model for ZrO_2 in section 3.1. The local charge neutrality of the TBBM is clearly inadequate to describe an ionic crystal for which a dominant part of the total energy is the Madelung sum of electrostatic pair terms.[10] A way to deal with this in tight binding was proposed by Majewski and Vogl[42,59] based on a Hubbard-like hamiltonian of Kittler and Falicov.[60] In this scheme the total charge transfer at each site, $\Delta q_{\mathbf{R}}$, from (11) and (19) are taken as point charges. The hamiltonian is again

$$H = H^0 + H' \tag{36}$$

as in (22). Two terms make up H', the Madelung energy of the lattice of point charges and a positive energy that is quadratic in $\Delta q_{\mathbf{R}}$, namely $U_{\mathbf{R}} \Delta q_{\mathbf{R}}^2$; employing the well-known "Hubbard U" that acts to resist the accumulation of charge. This problem is solved self consistently. An extension of this scheme to allow the charge to be expressed as multipoles, not just monopoles, was proposed independently by Schelling *et al.*[61] and Finnis *et al.*[38] In the latter paper, the connection was made to density functional theory and the TBBM, so we will pursue the same argument here. As noticed by Elstner *et al.*[62] the Hohenberg–Kohn total energy in DFT can be expanded about some reference electron density, $\rho^0(\mathbf{r})$. If H^0 is the hamiltonian with effective potential generated by the reference density, and just as in section 2.1 its eigenfunctions are $|n\rangle$ then the total energy correct to second order is[63] (e is the electron charge)

$$
\begin{aligned}
E^{(2)} &= \sum_n f_n \left\langle n \left| H^0 \right| n \right\rangle \\
&- \int \rho^0(\mathbf{r}) V^0_{\text{xc}}(\mathbf{r}) \mathrm{d}\mathbf{r} - E^0_{\text{H}} + E^0_{\text{xc}} + E_{\text{ZZ}} \\
&+ \frac{1}{2} \int \mathrm{d}\mathbf{r} \int \mathrm{d}\mathbf{r}' \left\{ e^2 \, \frac{\delta\rho(\mathbf{r}) \delta\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \right. \\
&+ \left. \delta\rho(\mathbf{r}) \frac{\delta^2 E_{\text{xc}}}{\delta\rho(\mathbf{r}) \delta\rho(\mathbf{r}')} \delta\rho(\mathbf{r}') \right\}. \qquad (37)
\end{aligned}
$$

E^0_{H} is the Hartree energy and E^0_{xc} and V^0_{xc} the exchange–correlation energy and potential belonging to the reference density, $\rho^0(\mathbf{r})$. The first two lines make up the Harris–Foulkes *first order* functional; we recognise the first line as the band energy, in fact the sum of occupied eigenvalues of the non self consistent *input* hamiltonian, and the second as the interaction term (double counting) plus the ion–ion pair potential, E_{ZZ}. In the *self consistent polarisable ion tight binding model*[38] (SCTB) we approximate the last two lines by a generalised Madelung energy and a Hubbard energy, which adds a *second order* energy[5] to (35)

$$
E_2 = \frac{1}{2} \, e^2 \sum_{\mathbf{R}L \, \mathbf{R}'L'} Q_{\mathbf{R}'L'} \, \tilde{B}_{\mathbf{R}'L' \, \mathbf{R}L} \, Q_{\mathbf{R}L} + \frac{1}{2} \sum_R U_{\mathbf{R}} \Delta q^2_{\mathbf{R}}. \qquad (38)
$$

These two terms represent the electron–electron interactions. All the exchange and correlation complexities are rolled into a single parameter, the Hubbard U. The first term in (38) is a classical interaction energy between point multipoles. The monopole term is just a straight forward sum of Coulomb energies, $\frac{1}{2} e^2 \, \Delta q_{\mathbf{R}} \Delta q_{\mathbf{R}'} / \, |\mathbf{R} - \mathbf{R}'|$, while the generalised Madelung matrix is just the LMTO bare structure constant matrix (28), or to be precise $B_{\mathbf{R}'L' \, \mathbf{R}L} = -(1/2\pi)(2\ell + 1)(2\ell' + 1) \tilde{B}_{\mathbf{R}'L' \, \mathbf{R}L}$. In general $Q_{\mathbf{R}L}$ is the multipole moment of angular momentum L at site $\mathbf{R}$. If we knew the charge density, which we don't in tight binding, then we could define the moment

$$
Q_{\mathbf{R}L} = \int \mathrm{d}\mathbf{r} \; \rho(\mathbf{r}) \; r^\ell \, Y_L(\mathbf{r}) \qquad (39)
$$

for $\ell > 0$; while for $\ell = 0$ we'll have

$$
Q_{\mathbf{R}0} = \Delta q_{\mathbf{R}} Y_0 = \sqrt{\frac{1}{4\pi}} \, \Delta q_{\mathbf{R}}.
$$

Although we don't know the charge density in tight binding, we know the eigenvectors of the hamiltonian and we can construct multipole moments from these. The monopole is of course proportional to the Mulliken charge transfer. Although in tight binding we don't even specify what the basis functions (3) are, we can take it that they comprise a radial part times an angular, spherical harmonic part, that is

$$
\langle \mathbf{r} | \mathbf{R}L \rangle = f_{\mathbf{R}\ell} \left(|\mathbf{r} - \mathbf{R}| \right) Y_L(\mathbf{r} - \mathbf{R}). \qquad (40)
$$

Then in terms of the eigenvector expansion coefficients (4), for $\ell > 0$ we may define

$$
Q_{\mathbf{R}L} = \sum_{L'L''} \sum_n f_n \, \bar{c}^n_{\mathbf{R}L'} c^n_{\mathbf{R}L''} \left\langle \mathbf{R}L' \left| \hat{Q}_{\mathbf{R}L} \right| \mathbf{R}L'' \right\rangle \qquad (41)
$$

in which the multipole moment operator is[64]

$$\hat{Q}_{\mathbf{R}L} = \hat{r}^\ell \, Y_L(\hat{\mathbf{r}}), \tag{42}$$

which follows as a consequence of (39). If we expand out the matrix element of $\hat{Q}_{\mathbf{R}L}$ using (40) and (42) we have

$$\begin{aligned}\left\langle \mathbf{R}L' \left| \hat{Q}_{\mathbf{R}L} \right| \mathbf{R}L'' \right\rangle &= \int r^2 \mathrm{d}r \; f_{R\ell'} \, f_{R\ell''} \, r^\ell \iint \mathrm{d}\Omega \, Y_{L''} \, Y_{L'} \, Y_L \\ &= \Delta_{\ell'\ell''\ell} \, C_{L'L''L},\end{aligned}$$

which introduces new tight binding parameters, $\Delta_{\ell'\ell''\ell}$. Selection rules which are policed by the Gaunt coefficients (29) demand that there are only seven new parameters, or two if one has a basis of only s and p orbitals. These parameters are

$$\begin{aligned}\Delta_{011} = \Delta_{101} &= \Delta_{spp} \\ \Delta_{112} &= \Delta_{ppd} \\ \Delta_{022} = \Delta_{202} &= \Delta_{sdd} \\ \Delta_{121} = \Delta_{211} &= \Delta_{pdp} \\ \Delta_{222} &= \Delta_{ddd} \\ \Delta_{123} = \Delta_{213} &= \Delta_{pdf} \\ \Delta_{224} &= \Delta_{ddg}.\end{aligned}$$

In fact these parameters are not entirely new, but are recognisable as the elements of crystal field theory—in the case $\ell' = \ell''$ they are the quantities $\langle r^\ell \rangle$.[65,66] So it's perhaps not surprising that these new parameters introduce *crystal field* terms into the hamiltonian. These are off-diagonal, on-site terms that we have up to now taken to be zero. However they are crucial in describing the bands of, for example, the transition metal oxides as in figure 3. The generalised Madelung energy in (38) implies that the electrons are seeing an electrostatic potential due to the multipole moments at all the atomic sites. Indeed, if the electrostatic potential in the neighbourhood of the atom at site $\mathbf{R}$ is expanded into spherical waves, we could write,

$$V_{\mathbf{R}}(\mathbf{r}) = \sum_L V_{\mathbf{R}L} \, r^\ell \, Y_L(\mathbf{r}) \tag{43}$$

and using standard electrostatics the $\mathbf{R}L$ coefficient in this expansion is

$$V_{\mathbf{R}L} = \sum_{\mathbf{R}'L'} \tilde{B}_{\mathbf{R}L\,\mathbf{R}'L'} \, Q_{\mathbf{R}'L'}.$$

Now in the same way that we arrived at (41), using (43) we can find the matrix elements of H', namely

$$H'_{\mathbf{R}L'\,\mathbf{R}L''} = U_{\mathbf{R}} \, \Delta q_{\mathbf{R}} \, \delta_{L'L''} + e^2 \sum_L V_{\mathbf{R}L} \, \Delta_{\ell'\ell''\ell} \, C_{L'L''L}. \tag{44}$$

Now all the ingredients of the self consistent tight binding scheme are assembled. H^0 is given by its matrix elements, determined as in non self consistent tight binding, described in section 3. After solving the orthogonal, or non orthogonal eigenproblem and finding

the eigenvector expansion coefficients, you build the multipole moments and using structure constants find the components, $V_{\mathbf{R}L}$, of the potential. Having also chosen the Δ and Hubbard U parameters, elements of H' are assembled and the eigenproblem is solved for $H^0 + H'$. This continues until self consistency.

One or two extensions have been omitted here.

1. Only *on-site* matrix elements of H' are non zero in this self consistent scheme. In fact in the case of a non orthogonal basis, due to the explicit appearance of bond charge (see equation (11) and subsequent remarks) also intersite matrix elements of H' are introduced. This is important because it allows the hopping integrals themselves to be affected by the redistribution of charge, as might be intuitively expected.[6,67] Details are to be found elsewhere.[5,23]

2. This scheme can be extended to admit spin polarisation in imitation of the local spin density approximation. This *magnetic tight binding* (figure 1) has also been described elsewhere and is omitted from these notes for brevity.[23]

Finally we should remark that the interatomic force is easily obtained in self consistent tight binding. Only the *first* and *third* terms in the TBBM (21) survive; in particular one still requires the derivatives of the matrix elements of H^0. The only additional contribution to the force comes from the *first term* in (38); there is no contribution from the second term (or from the Stoner term in magnetic tight binding[68]) because of the variational principle. Hence one requires only the classical electrostatic force on atom $\mathbf{R}$,

$$\mathbf{F}^{\text{es}}_{\mathbf{R}} = -\sum_{L} Q_{\mathbf{R}L} \boldsymbol{\nabla} V_{\mathbf{R}L}$$

which is consistent with the force theorem,[31–34] and repairs the inconsistency of the band model mentioned in section 2.3.

We illustrated the self consistent polarisable ion tight binding model (SCTB) in the study of phase transitions in ZrO_2 in section 3.1. It turns out that the extension of the point charge model to include polarisability introduces new physics that is essential in describing these phenomena. In particular the dipole polarisation of the anions drives the cubic to tetragonal transition. Furthermore, as seen in figure 3 the crystal field splitting of the cation d-bands is achieved naturally and the correct ordering is reproduced in cubic and octahedral crystal fields. Crystal field splitting is also largely responsible for the ligand bandwidth in the low symmetry rutile structure.

4.1 Application to small molecules

Now we will turn to a second example, the application to small molecules. The self consistent point charge model in this context and in the study of biological molecules has enjoyed enormous success thanks in particular to the work of Frauenheim, Elstner and colleagues.[69]

Here we demonstrate the SCTB model applied to the question of the polarisability of two small molecules, azulene and para-nitroaniline (pNA). Hopping parameters were taken from Horsfield *et al.*[70] and Hubbard U and Δ parameters chosen to to reproduce the ground state dipole moments predicted by the local density approximation. For azulene it is found that the self consistent point charge model is sufficient, but pNA cannot be described

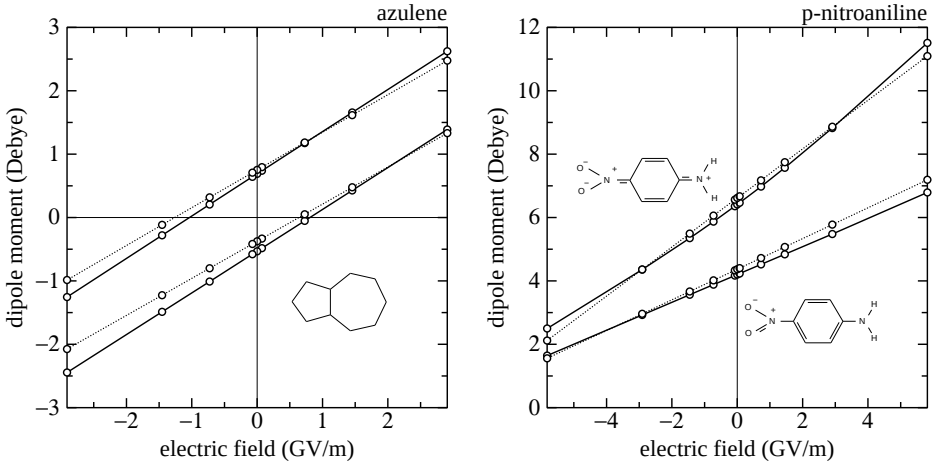


Figure 6. Dipole moment as a function of applied electric field calculated using LSDA, solid lines, and SCTB, dotted lines.[71] LSDA calculations were made using a molecule LMTO program.[72,73] The left hand figure shows the molecule azulene and the upper set of lines refer to the ground state and lower set to the so called S_1 excited state. The right hand figure shows p-nitroaniline; the lower set are the ground state and the upper set the "zwitterionic" first excited state

properly without dipole polarisability.[71] Figure 6 shows that the SCTB model provides a very accurate rendering of the dipole response to an applied electric field compared to LSDA calculations. We discuss now the two molecules in turn.

1. Azulene is a very interesting molecule having the same chemical formula as napthalene but comprising a five and seven membered ring instead of two six membered rings. According to Hückel's "$4n + 2$ rule," a ring molecule is especially stable if it has N π-electrons and $N = 4n + 2$, where n is an integer. This is because this leads to a closed shell of π-electrons.[74] Hence benzene is stable, having $n = 1$. By a similar argument a seven membered ring has an unpaired electron which can be used to occupy an unpaired hole in a five membered ring. Hence the ground state of azulene possesses a large dipole moment. An excited state is created if the electron is returned to the seven membered ring. As shown to the left of figure 6 the ground state dipole moment is positive (the positive axis pointing to the right) while its sign is reversed in the first excited state. Here we use a device which is not quite legitimate, namely in both LSDA and SCTB an electron–hole pair is created and self consistency arrived at under this constraint. While a very crude approximation to an excited state[75] (given that LSDA is a ground state theory) this does provide a useful test of the validity of the SCTB model. Indeed it is quite remarkable how the SCTB faithfully reproduces the LSDA even to the extent of accurately reproducing the polarisability of both ground and excited states. (The polarisability is the linear response of the dipole moment to an applied electric field, namely the slope in these figures.)

2. pNA is the archetypal "push–pull" chromophore.[76] In the ground state the dipole moment is small, but the first excited state is thought to be "zwitterionic," meaning

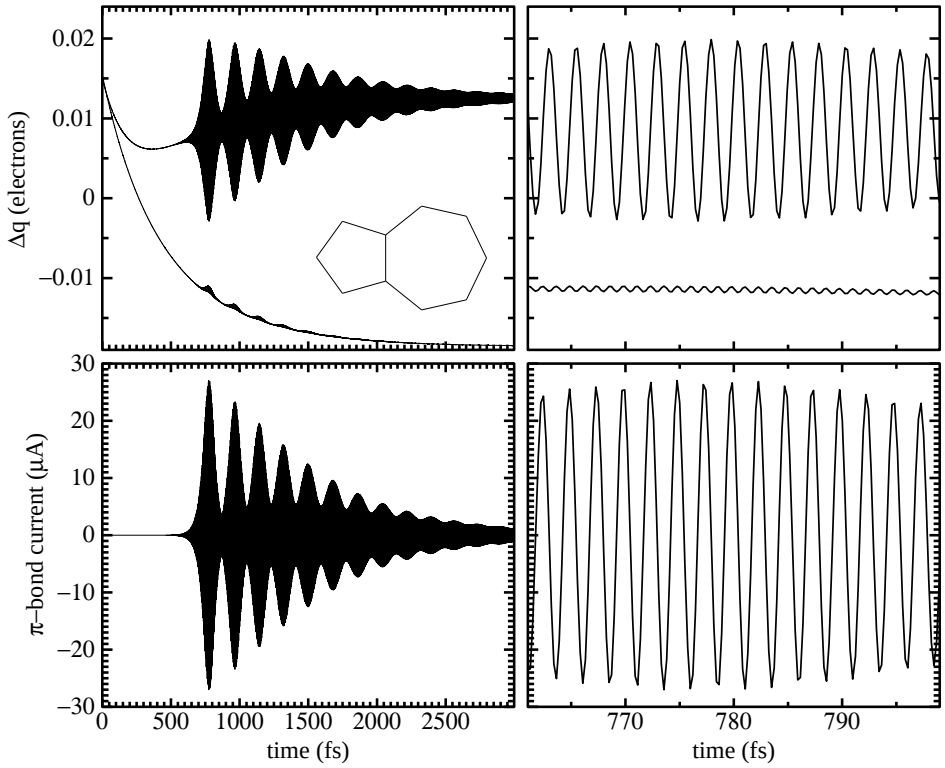


Figure 7. Charge transfer and bond current as a function of time in the relaxation of the S_2 excited state in azulene. The upper panels show the excess charge on a "bridge" atom and on the rightmost atom in the seven membered ring (lower curve). The lower panels show the π–bond current in the "bridge" bond.

that an electron transfers from the amine group on the right to the NO_2 group at the left increasing the dipole moment as shown on the right hand side of figure 6. Transfer of the electron through the π-system is called a push–pull process. Again the SCTB faithfully reproduces the LSDA with quantitative accuracy. We should mention again that it did not seem possible to obtain this result using a point charge self consistent tight binding model.

4.2 Ring currents in azulene

The SCTB model provides a simple scheme for the study of electron transfer as in the push–pull process. This is done by solving the time dependent Schrödinger equation using the hamiltonian H including electron–electron interactions. Indeed this is probably the simplest quantum mechanical model that goes beyond non interacting electrons. We have applied this approach to the relaxation of the S_2 excited state in azulene with some quite spectacular results.[77]

In terms of the density operator, the time dependent Schrödinger equation is

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\hat{\rho} = (\mathrm{i}\hbar)^{-1}\,[H,\hat{\rho}] - \Gamma\,(\hat{\rho} - \hat{\rho}_0)\,.$$

We have added a damping term with time constant Γ^{-1}. This allows us to prepare the molecule in an excited state and relax it into the ground state whose density operator is $\hat{\rho}_0$. The equation of motion is solved numerically using a simple leapfrog algorithm. While at the outset, the density matrix is real, during the dynamics it acquires complex matrix elements whose imaginary parts describe *bond currents*,[18]

$$j_{\mathbf{R}\mathbf{R}'} = \frac{2e}{\hbar}\sum_{LL'} H_{\mathbf{R}'L'\,\mathbf{R}L}\ \mathrm{Im}\,\rho_{\mathbf{R}L\,\mathbf{R}'L'}$$

which is the total current flowing from atom $\mathbf{R}$ to atom $\mathbf{R}'$. By selecting certain L-channels we can extract orbital contributions to j; in the present case of push–pull transfer we are interested in the current carried by the π-system of electrons.

Figure 7 shows results of such a simulation in azulene, using a time constant $\Gamma^{-1} = 500$ fs. Examine first the lower curve in the upper left panel. This is the excess total charge on the rightmost atom in the seven membered ring (see the inset in the top left panel). In the excited state, the dipole moment points to the left, that is, there is excess charge on this atom which transfers through the π-system to the left as the molecule relaxes into the ground state for which the dipole moment has opposite sign. The curve clearly show a smooth transfer of charge away from this site. However superimposed upon this is a series of oscillatory excursions in charge transfer, shown in a narrow time window by the lower curve in the upper right panel. Accompanying these oscillations are much larger fluctuations in the charge on the upper atom belonging to the "bridge" bond which is shared by both the five and seven membered rings. This excess charge is plotted in the upper curves of the upper left and right hand panels. As the upper and lower left hand panels show these oscillations die away, but analysis shows a quite characteristic frequency as seen in the right hand panels. The lower two panels show the π-bond current in the "bridge" bond. What is happening here is the setting up of ring currents in both rings whose directions are alternating with a period of a few femtoseconds. The ring currents at any one time are travelling in opposite senses in the two rings. This phenomena is a consequence of the electron–electron interaction, as we can verify by repeating the calculations using the non interacting hamiltonian, H^0. Because two bonds enter each bridge atom but only one leaves, the opposing sense of the currents means that charge will accumulate on one of these atoms to the point at which the Coulomb repulsion (described by the Hubbard U) resists further current flow and indeed reverses its direction. Note that each current reversal (lower right panel) is mirrored by the alternating charge transfer on the bridge atoms (upper right panel). It is not yet understood what fixes the frequency at which the reversal happens or what it is that makes the molecule particularly susceptible to this instability. We note that these ring currents require a long lead-in time, on the order of the time constant, to become established and this is probably because the symmetry breaking comes about through numerical round-off in the computer. In a more detailed simulation coupling the electrons to the molecular vibrations,[78] this symmetry breaking will derive from the coupling. We can confirm that the great majority of the current is indeed carried by the π-electron system.

5 Last Word

The intention here has been to provide a practical introduction to the tight binding method and to motivate students to try it for themselves. While this is a long article it is mostly conspicuous for what is missing, rather than what is included. This is not surprising in view of the vast literature and considerable age of the tight binding approximation, but I've tried to bring out issues that are less widely discussed elsewhere. Regrettably no connection has been made to the semi empirical approaches in quantum chemistry that bear a close resemblance. This reflects the fact that physicists and chemists frequently discover the same science independently and often without much awareness of each other's work. I hope that some of the most glaring omissions will be covered by other authors in this volume.[8,79]

Appendix

Real spherical harmonics are described in ref [64]. One takes the conventional, complex spherical harmonics[80] and makes linear combinations to get the real and imaginary parts.[81] Instead of m running from $-\ell$ to ℓ, m now runs from 0 to ℓ but for each $m > 0$, there are two real functions: $Y^c_{\ell m}$ which is $(-1)^m\sqrt{2}$ times the real part of $Y_{\ell m}$; and $Y^s_{\ell m}$ which is $(-1)^m\sqrt{2}$ times the imaginary part of $Y_{\ell m}$. For $m = 0$, $Y_{\ell m}$ is anyway real, so we throw away $Y^s_{\ell 0}$. We end up with the same number of functions, properly orthonormal. Specifically,

$$Y^c_{\ell m} = (-1)^m \frac{1}{\sqrt{2}} \left(Y_{\ell m} + \bar{Y}_{\ell m}\right)$$
$$Y^s_{\ell m} = (-1)^m \frac{1}{\mathrm{i}\sqrt{2}} \left(Y_{\ell m} - \bar{Y}_{\ell m}\right).$$

References

1. J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498, 1954
2. W. A. Harrison, *Electronic structure and the properties of solids: the physics of the chemical bond*, (W. H .Freeman, San Fransisco, 1980)
3. A. P. Sutton, *Electronic structure of materials*, (Clarendon Press, Oxford, 1993). The Slater–Koster table is reproduced in table 9.3
4. D. G. Pettifor, *Bonding and structure in molecules and solids*, (Clarendon Press, Oxford, 1995)
5. M. W. Finnis, *Interatomic forces in condensed matter*, (Oxford University Press, 2003)
6. A. P. Sutton, M. W. Finnis, D. G. Pettifor and Y. Ohta, J. Phys. C: Solid State Phys. **21**, 35, 1988
7. W. M. C. Foulkes, Phys. Rev. B **48**, 14216, 1993
8. R. Drautz, in this volume
9. O. K. Andersen, O. Jepsen and D. Glötzel, *Canonical description of the band structures of metals*, in *Proc. Intl. Sch. Phys., LXXXIX Corso, Varenna*, edited by F. Bassani, F. Fumi and M. P. Tosi (Soc. Ital. di Fisica, Bologna, 1984) p. 59

10. N. W. Ashcroft and N. D. Mermin, *Solid state physics*, (Holt-Saunders, 1976)
11. O. Jepsen and O. K. Andersen, Solid State Commun. **9**, 1763, 1971
12. G. Kresse and J. Furthmüller, Comp. Mater. Sci. **6**, 15, 1996
13. M. Reese, M. Mrovec and C. Elsässer, to be published
14. S. Glanville, A. T. Paxton and M. W. Finnis, J. Phys. F: Met. Phys. **18**, 693, 1988
15. A. T. Paxton, Phil. Mag. B **58**, 603, 1988
16. A. P. Horsfield, private communications
17. A. P. Horsfield, A. M. Bratkovsky, M. Fearn, D. G. Pettifor and M. Aoki, Phys. Rev. B **53**, 12694, 1996
18. T. N. Todorov, J. Phys.: Condens. Matter **14**, 3049, 2002
19. L. I. Schiff, *Quantum mechanics, 3*$^{\mathrm{rd}}$ *ed.*, (McGraw-Hill, 1968) p. 379
20. C. Kittel, *Quantum theory of solids, 2*$^{\mathrm{nd}}$ *ed.*, (John Wiley, 1987) p. 101
21. L. E. Ballentine and M. Kolář, J. Phys. C: Solid State Phys. **19**, 981, 1986
22. S. Elliott, *The physics and chemistry of solids*, (Wiley, 1998)
23. A. T. Paxton and M. W. Finnis, Phys. Rev. B **77**, 024428, 2008
24. J. Friedel, Suppl. Nuovo Cimento, Serie X **VII**, 287, 1958
25. P. W. Anderson, Phys. Rev. **124**, 41, 1961
26. M. Aoki, Phys. Rev. Letters **71**, 3842, 1993
27. D. J. Chadi, Phys. Rev. Letters **41**, 1062, 1978
28. J. Friedel, Trans. Metall. Soc. AIME **230**, 616, 1964
29. F. Ducastelle, J. de Physique **31**, 1055, 1970
30. D. G. Pettifor, in *Physical Metallurgy, 3*$^{\mathrm{rd}}$ *ed.*, edited by R. W. Cahn and P. Hassen, (Elsevier, 1983)
31. D. G. Pettifor, J. Chem. Phys. **69**, 2930, 1978
32. M. Methfessel and J. Kübler, J. Phys. F: Met. Phys. **12**, 141, 1982
33. A. R. Mackintosh and O. K. Andersen, in *Electrons at the Fermi surface*, edited by M. Springford (Cambridge University Press, 1980) p. 149
34. V. Heine, Solid State Physics **35**, edited by H. Ehrenreich, F. Seitz and D. Turnbull, (Academic Press, 1980) p. 1
35. D. Spanjaard and M. C. Desjonquères, Phys. Rev. B **30**, 4822, 1984
36. A. T. Paxton, J. Phys. D: Appl. Phys. **29**, 1689, 1996
37. H. Haas, C. Z. Wang, M. Fähnle, C. Elsässer and K. M. Ho, Phys. Rev. B **57**, 1461, 1998
38. M. W. Finnis, A. T. Paxton, M. Methfessel and M. van Schilfgaarde, Phys. Rev. Letters **81**, 5149, 1998
39. D. W. Richerson, *Modern ceramic engineering*, (Marcel Dekker, New York, 1992)
40. S. Fabris, A. T. Paxton and M. W. Finnis, Phys. Rev. B **63**, 94101, 2001
41. S. Fabris, A. T. Paxton and M. W. Finnis, Acta Materialia **50**, 5171, 2002
42. J. A. Majewski and P. Vogl, in *The structures of binary compounds*, edited by F. R. de Boer and D. G. Pettifor, (Elsevier Science Publishers, 1989) p. 287
43. O. K. Andersen, Solid State Commun. **13**, 133, 1973
44. D. G. Pettifor, J. Phys. F: Met. Phys. **7**, 613, 1977
45. `http://titus.phy.qub.ac.uk/Programs`
46. L. Goodwin, A. J. Skinner and D. G. Pettifor, Europhys. Letters **9**, 701, 1989
47. D. G. Pettifor, J. Phys. C: Solid State Phys. **5**, 97, 1972
48. O. K. Andersen, O. Jepsen and M. Šob, in Lecture Notes in Physics **282**, *Electronic*

band structure and its applications, edited by M. Yusouff, (Springer, Berlin, 1987)
49. O. K. Andersen and O. Jepsen, Phys. Rev. Letters **53**, 2571, 1984
50. A. R. Williams, J. Kübler and C. D. Gelatt, Jr., Phys. Rev. B **19**, 6094, 1979
51. H. Nakamura, D. Nguyen-Manh and D. G. Pettifor, J. Alloys Compounds **306**, 113, 2000
52. R. W. Tank and C. Arcangeli, phys. stat. sol. (b) **217**, 131, 2000
53. D. Nguyen-Manh, D. G. Pettifor and V. Vitek, Phys. Rev. Letters **85**, 4136, 2000
54. A. T. Paxton, in *Atomistic simulation of materials: beyond pair potentials*, edited by V. Vitek and D. J. Srolovitz, (Plenum, 1989) p. 327
55. M. S. Tang, C. Z. Wang, C. T. Chan and K. M. Ho, Phys. Rev. B **53**, 979, 1996
56. M. Mrovec, D. Nguyen-Manh, D. G. Pettifor and V. Vitek, Phys. Rev. B **69**, 094115, 2004
57. M. Mrovec, R. Gröger, A. G. Bailey, D. Nguyen-Manh, C. Elsässer and V. Vitek, Phys. Rev. B **75**, 104119, 2007
58. A. J. Skinner and D. G. Pettifor, J. Phys.: Condens. Matter **3**, 2029, 1991
59. J. A. Majewski and P. Vogl, Phys. Rev. Letters **57**, 1366, 1986
60. R. C. Kittler and L. M. Falicov, Phys. Rev. B **18**, 2506, 1978
61. P. K. Schelling, N. Yu and J. W. Halley, Phys. Rev. B **58**, 1279, 1998
62. M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai and G. Seifert, Phys. Rev. B **58**, 7260, 1998
63. W. M. C. Foulkes and R. Haydock, Phys. Rev. B **39**, 12520, 1989
64. A. J. Stone, *The theory of intermolecular forces*, (Oxford University Press, 1996)
65. D. S. McClure, in *Phonons*, edited by R. W. H. Stevenson, (Oliver and Boyd, London, 1966) p. 314
66. A. M. Stoneham, *Theory of defects in solids*, (Oxford University Press, 1975)
67. M. Lannoo and J. Bourgoin, *Point defects in semiconductors I*, Springer Series in Solid State Sciences, **22**, (Springer, Berlin, 1981)
68. G. Liu, D. Nguyen-Manh, B.-G. Liu and D. G. Pettifor, Phys. Rev. B **71**, 174115, 2005
69. M. Elstner, Th. Frauenheim, J. McKelvey and G. Seifert, in *Special Section: DFTB symposium*, J. Phys. Chem. A **111**, 5607–5944, 2007
70. A. P. Horsfield, P. D. Godwin, D. G. Pettifor and A. P. Sutton, Phys. Rev. B **54**, 15773, 1996
71. A. T. Paxton, unpublished
72. M. Methfessel and M. van Schilfgaarde, Phys. Rev. B **48**, 4937, 1993
73. A. T. Paxton and J. B. Harper, Mol. Phys. **102**, 953, 2004
74. R. McWeeny, *Coulson's valence*, (Oxford University Press, 1979) p. 243
75. A. Hinchliffe and H. J. Soscún, Chem. Phys. Letters **412**, 365, 2005
76. A. M. Moran and A. Myers Kelly, J. Chem. Phys. **115**, 912, 2001
77. A. M. Elena, A. T. Paxton and T. N. Todorov, to be published
78. A. P. Horsfield, D. R. Bowler, H. Ness, C. G. Sánchez, T. N. Todorov and A. J. Fisher, Rep. Prog. Phys. **69**, 1195, 2006
79. M. Elstner, in this volume
80. J. D. Jackson, *Classical electrodynamics, 2*$^{\rm nd}$ *ed.*, (John Wiley, 1975) p. 99
81. M. Methfessel, *Multipole Green functions for electronic structure calculations*, (Katholieke Universiteit te Nijmegen, 1986)

Two Topics in Ab Initio Molecular Dynamics: Multiple Length Scales and Exploration of Free-Energy Surfaces

Mark E. Tuckerman

Department of Chemistry and
Courant Institute of Mathematical Sciences
100 Washington Square East
New York University, New York, NY 10003
E-mail: mark.tuckerman@nyu.edu

This lecture will consider two problems that arise in molecular dynamics simulations of complex systems. The first is the treatment of multiple length scales in simulations that employ reciprocal-space techniques (plane-wave basis sets in *ab initio* molecular dynamics, Ewald summation,...) for the calculation of long-range forces. It will be shown that a dual-gridding scheme, with reciprocal space grids of two very different resolutions, can be used to substantially reduce the cost of the calculation without sacrificing accuracy. Interpolation between the two grids is achieved via the use of Euler exponential splines commonly employed in the particle-mesh Ewald method. Two application areas from *ab initio* molecular dynamics will be illustrated, namely, the use of dual-gridding in QM/MM calculations and in cluster calculations with plane-wave basis sets. The second problem is an inherently multiple time-scale problem involving the exploration of rough free-energy surfaces. It will be shown that efficient exploration and calculation of such surfaces is possible using an adiabatic dynamics technique in which a subset of collective variables are "driven" at high temperature by a set of external driving variables whose masses are adjusted so as to effect an adiabatic decoupling of the collective variables from the remainder of the system. Under these conditions, free-energy surfaces can be constructed straightforwardly from the probability distribution functions generated with the requirement of adjusting only a few parameters. The method will be illustrated on the folding of an alanine hexamer.

1 The Multiple Length-Scale Problem

Chemical systems are often characterized by a set of electronically active constituents localized in a small region of space, surrounded by and interacting with a large bath of electronically inert components. The division of a large system into chemically interesting and chemically uninteresting regions is both an intuitively appealing and a practically useful description that can be fruitfully applied to many different problems of chemical and biological interest. For example, in the study of solution phase chemical reactions, it is advantageous to consider, explicitly, the electronic degrees of freedom of the reactants and products and perhaps a first solvation shell, while the remainder of the solvent is modeled more approximately[1]. Similarly, in studies of enzyme catalysis, the valence electrons of the amino acids and the water molecules near the active site, as well as those of the substrate, must be modeled using a high level of theory while the remainder of these large and complex systems can be modeled more approximately[1–6]. Thus, simulation studies based on hybrid model descriptions promise to yield chemical insight into significant problems for low computational cost. It is, therefore, important to develop both the models and methods required to treat mixed *ab initio*/empirical force field descriptions of chemical and biological systems accurately and efficiently. In addition, systems with reduced peri-

odicity that require a vacuum region in the non-periodic direction, require methodological developments to reduce the inefficiency of treating a large empty spatial region.

It has been demonstrated that a wide variety of complex chemical systems can be treated effectively using an *ab initio* methodology that employs a plane wave basis set in conjunction with the generalized gradient approximation to density functional theory (GGA-DFT)[7–10]. Of course, in realistic calculations, many basis functions or equivalently, a large plane wave energy cutoff ($E_{cut} = \hbar^2 g_{max}^2/2m_e$) must be employed to ensure accuracy. The large basis set size coupled with the fact that plane waves, naturally, admit, only a single length scale has made it difficult to employ plane wave based GGA-DFT to study hybrid model systems.

Consider, for example, a small simulation cell containing the electron density embedded within a large simulation cell containing the rest of the system (i.e. the bath) or possibly empty space. In order to determine the long range interaction of the electron density with the atoms outside the small simulation cell within the plane wave formalism, it is necessary to expand the electron density in the large simulation cell using the **same** large cutoff required to describe the rapidly varying electron density in the small cell (e.g. $E_{cut} \approx 70$ Ry). Thus, the memory requirements are prohibitively large and the calculations scale poorly with the size of the large cell (at fixed small cell size). However, such a scheme does allow systems modeled using 3D periodic boundary conditions (liquids and solids) to be accurately studied. It also permits novel reciprocal space based techniques that treat clusters, wires and surfaces appropriately[11–13] (open, and 1D and 2D periodic boundary conditions, respectively), properly, to be applied to "mixed" or "hybrid" model calculations.

In this lecture, we will describe a dual-gridding method[14] designed to treat large systems that can be decomposed into electronically active and electronically inert portions with high efficiency is presented. Two length scales are explicitly introduced into the plane wave based GGA-DFT electronic structure formalism so that the small length scale, electronically active region can be treated differently than the long length scale, electronically inert region without loss of generality and with large gains in scalability and efficiency. This is accomplished by employing a Cardinal B-spline based formalism to derive a novel expression for the electronic energy that explicitly contains both the long and short length scales. The new expression can be evaluated efficiently using two independent plane wave energy cutoffs and is smooth, differentiable, and rapidly convergent with respect to the plane wave cutoff associated with the long length scale even when the plane wave cutoff associated with the short length scale is quite large. Thus, the method scales as $N \log N$ where N is number of atoms in the full system (at fixed size of the chemically active region) provided particle mesh Ewald techniques[15–18] are employed to evaluate the atomic charge density in the large cell. In addition, the new methodology does not involve an *ad hoc* electrostatic potential fitting scheme based on point charges derived from a particular choice of population analysis and can be utilized to treat clusters, wires, surfaces and solids/liquids without loss of generality. We note that a similar approach was recently developed for use in Gaussian-based QM/MM calculations[19].

1.1 Methods

In the Kohn-Sham formulation of density functional theory, the electron density is expanded in a set of orbitals $\{\psi_i(\mathbf{r})\}$

$$n(\mathbf{r}) = \sum_{i=1}^{n} |\psi_i(\mathbf{r})|^2 \tag{1}$$

and the energy functional is given by

$$E[n] = T_s[\{\psi_i\}] + E_{\rm H}[n] + E_{\rm xc}[n] + E_{\rm ext}[n] \tag{2}$$

where T_s is the kinetic energy of a system of noninteracting electrons, $E_{\rm H}$ is the Hartree energy, and $E_{\rm xc}$ is the exchange and correlation energy. For example, one could employ a Generalized Gradient Approximation such as the BLYP (Becke86 exchange and LYP correlation) functional[20,21]. In this case, Eq. (2) is referred to as a GGA-density functional.

In this work, the GGA-density functional, Eq. (2), is minimized by expanding the orbitals in a finite plane wave basis set and varying the expansion coefficients subject to the orthogonality constraints ($\langle\psi_j|\psi_i\rangle = \delta_{ij}$). The plane wave basis set is truncated by including all plane waves with kinetic energy less than or equal to a cutoff energy, $\hbar g^2/2m_e \leq E_{cut}$. Finally, core electrons, which are difficult to treat in a plane-wave basis set, are replaced by atomic pseudopotentials. Typical pseudopotentials contain a long range local contribution, $E_{\rm loc}$, and a short range angular momentum dependent nonlocal contribution, $E_{\rm nonloc}$ that serves to replace the core (i.e. $E_{\rm ext} = E_{\rm loc} + E_{\rm nonloc}$).

The GGA-density functional, therefore, contains only two terms that act at long range, specifically, the Hartree, $E_{\rm H}[n]$, and local pseudopotential, $E_{\rm loc}[n]$, energies defined by

$$E_{\rm H}[n] = \frac{e^2}{2} \sum_{\hat{\mathbf{S}}} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}' \, \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}' + \overleftrightarrow{\mathbf{h}}\hat{\mathbf{S}}|} \tag{3}$$

$$E_{\rm loc}[n] = \sum_{\hat{\mathbf{S}}} \sum_{I=1}^{N} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \, \phi_{{\rm loc},I}(\mathbf{r} - \mathbf{R}_I + \overleftrightarrow{\mathbf{h}}\hat{\mathbf{S}}) n(\mathbf{r}) \tag{4}$$

where $\mathbf{R}_I$ is the Cartesian position of the Ith ion, $\overleftrightarrow{\mathbf{h}}$ is the cell matrix whose columns contain the d cell vectors, $\det \overleftrightarrow{\mathbf{h}} = V$ is the volume, and $\hat{\mathbf{S}} = \{\hat{s}_a, \hat{s}_b, \hat{s}_c\}$ is a vector of integers indexing the periodic replicas (in clusters, only $\hat{\mathbf{S}} = \{0, 0, 0\}$ is allowed while in systems periodically replicated in three spatial dimensions, the three integers span the full range).

In plane-wave based calculations the orbitals and, the density are expanded as follows,[8,22]

$$\psi_j(\mathbf{r}) = \frac{1}{\sqrt{V}} \sum_{\hat{\mathbf{g}}} \bar{\psi}_j(\mathbf{g}) \exp(i\mathbf{g}\cdot\mathbf{r})$$

$$n(\mathbf{r}) = \frac{1}{V} \sum_{\hat{\mathbf{g}}} \bar{n}(\mathbf{g}) \exp(i\mathbf{g}\cdot\mathbf{r}), \tag{5}$$

where $\mathbf{g} = \overleftrightarrow{\mathbf{h}}^{-1} \hat{\mathbf{g}}$ and the vector of integers, $\hat{\mathbf{g}} = \{g_a, g_b, g_c\}$, indexes reciprocal space. Typically, a cutoff is introduced on the sums describing the orbitals such that $\hbar g^2/2m_e \leq$

E_{cut}. (Note, the reciprocal space summation for the density is over a reciprocal space defined by the appropriately larger cutoff, $E_{cut}^{(\mathrm{density})} = 4E_{cut}$.) It is convenient to express the Hartree and local external energies in reciprocal space

$$E_{\mathrm{H}} = \frac{e^2}{2V} \sum_{\hat{\mathbf{g}}}{}' |\bar{n}(\mathbf{g})|^2 \left[\frac{4\pi}{g^2} + \hat{\phi}^{(\mathrm{screen,Coul})}(\mathbf{g}) \right] + \left(\frac{e^2}{2V} \right) \hat{\phi}^{(\mathrm{screen,Coul})}(0) |\bar{n}(0)|^2 \tag{6}$$

$$E_{\mathrm{loc}} = \frac{1}{V} \sum_{\hat{\mathbf{g}}}{}' \sum_{I=1}^{N} \bar{n}^*(\mathbf{g}) \exp(-i\mathbf{g}\cdot\mathbf{R}_I) \left[\tilde{\phi}_{\mathrm{loc},I}(\mathbf{g}) - eq_I \hat{\phi}^{(\mathrm{screen,Coul})}(\mathbf{g}) \right] + \frac{1}{V} \sum_{I=1}^{N} \bar{n}(0) \left[\tilde{\phi}^{(0)}_{\mathrm{loc},I} - eq_I \hat{\phi}^{(\mathrm{screen,Coul})}(0) \right]. \tag{7}$$

Here, $\tilde{\phi}_{\mathrm{loc},I}$ denotes the Fourier Transform of the local pseudopotential, the prime indicates that the $\mathbf{g} = 0$ term is eliminated and the function, $\tilde{\phi}^{(0)}_{\mathrm{loc},I}$ is the non-singular part of the local pseudopotential at $\mathbf{g} = 0$. The screening function, $\hat{\phi}^{(\mathrm{screen,Coul})}(\mathbf{g})$, is added to treat systems with fewer than three periodic dimensions (clusters, surfaces, wires), as discussed in Refs.[11–13]. For a system periodically replicated in three spatial dimensions, it is identically zero.

It is clear that the standard expressions for the Hartree and local external energies given in Eq. (6) and Eq. (7), respectively, only possesses a single length scale. A second length scale can be introduced by first rewriting the real space expressions for these two energies using the identity $\mathrm{erf}(\alpha r) + \mathrm{erfc}(\alpha r) = 1$,

$$\begin{aligned} E_{\mathrm{H}}[n] &= \left\{ \frac{e^2}{2} \sum_{\hat{\mathbf{S}}} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}' \, \frac{n(\mathbf{r}) n(\mathbf{r}') \mathrm{erfc}(\alpha |\mathbf{r} - \mathbf{r}' + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|)}{|\mathbf{r} - \mathbf{r}' + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|} \right\} \\ &+ \left\{ \frac{e^2}{2} \sum_{\hat{\mathbf{S}}} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}' \, \frac{n(\mathbf{r}) n(\mathbf{r}') \mathrm{erf}(\alpha |\mathbf{r} - \mathbf{r}' + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|)}{|\mathbf{r} - \mathbf{r}' + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|} \right\} \\ &= E_{\mathrm{H}}^{(\mathrm{short})}[n] + E_{\mathrm{H}}^{(\mathrm{long})}[n] \end{aligned} \tag{8}$$

$$\begin{aligned} E_{\mathrm{loc}}[n] &= \left\{ \sum_{\hat{\mathbf{S}}} \sum_{I=1}^{N} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \; n(\mathbf{r}) \left[\phi_{\mathrm{loc},I}(\mathbf{r} - \mathbf{R}_I + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}) + \frac{eq_I \mathrm{erf}(\alpha |\mathbf{r} - \mathbf{R}_I + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|)}{|\mathbf{r} - \mathbf{R}_I + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|} \right] \right\} \\ &- \left\{ \sum_{\hat{\mathbf{S}}} \sum_{I=1}^{N} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \; n(\mathbf{r}) \left[\frac{eq_I \mathrm{erf}(\alpha |\mathbf{r} - \mathbf{R}_I + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|)}{|\mathbf{r} - \mathbf{R}_I + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|} \right] \right\} \\ &= E_{\mathrm{loc}}^{(\mathrm{short})}[n] + E_{\mathrm{loc}}^{(\mathrm{long})}[n]. \end{aligned} \tag{9}$$

Here, the first term in the curly brackets in each equation is short range while the second term is long range. Note, both $\phi_{\mathrm{loc},I}(\mathbf{r})$ and $-eq_I \mathrm{erf}(\alpha r)/r$ approach $-eq_I/r$, asymptotically where q_I is the charge on I^{th} ion core. In the limit $\alpha V^{1/3} >> 1$, the sum over

images in the first term of each expression (i.e. the short range parts) can be truncated at the first or nearest image with exponentially small error.

In order to proceed, it will be assumed that the electrons are localized in a particular region of the large cell described by $\overleftrightarrow{\mathbf{h}}$ which can be enclosed in a small cell, described by $\overleftrightarrow{\mathbf{h}}_s$, centered at the point, $\mathbf{R}_c$. That is, the orbitals and, hence, electron density are taken to vanish on the surface of $\overleftrightarrow{\mathbf{h}}_s$. Furthermore, it is assumed, for simplicity, that the a_s, b_s and c_s axes of $\overleftrightarrow{\mathbf{h}}_s$ are parallel to the a, b and c axes of $\overleftrightarrow{\mathbf{h}}$ such that $\overleftrightarrow{\mathbf{h}}^{-1}\overleftrightarrow{\mathbf{h}}_s = \overleftrightarrow{\mathbf{D}}$, a diagonal matrix. Thus, we can define,

$$\begin{aligned} \psi_j(\mathbf{r}_s + \mathbf{R}_c) &= \psi_{j,s}(\mathbf{r}_s) \\ n(\mathbf{r}_s + \mathbf{R}_c) &= n_s(\mathbf{r}_s) \end{aligned} \tag{10}$$

where, the $\mathbf{r}_s$ span the small cell and can be expressed as $\mathbf{r}_s = \overleftrightarrow{\mathbf{h}}_s\mathbf{s}$ with $0 \le s_\alpha \le 1$ and, both, $\psi_j(\mathbf{r}) \equiv 0$ and $n(\mathbf{r}) \equiv 0$ for $\mathbf{r}_s = \mathbf{r} - \mathbf{R}_c$ outside the domain of $\overleftrightarrow{\mathbf{h}}_s$. The orbitals and the electron density can be expanded in a plane wave basis set that spans the small cell, only,

$$\begin{aligned} \psi_{j,s}(\mathbf{r}_s) &= \frac{1}{\sqrt{V}_s}\sum_{\hat{\mathbf{g}}_s} \bar{\psi}_{j,s}(\mathbf{g}_s)\exp(i\mathbf{g}_s\cdot\mathbf{r}_s) \\ n_s(\mathbf{r}_s) &= \frac{1}{V_s}\sum_{\hat{\mathbf{g}}_s} \bar{n}_s(\mathbf{g}_s)\exp(i\mathbf{g}_s\cdot\mathbf{r}_s) \quad , \end{aligned} \tag{11}$$

where $\mathbf{g}_s = \overleftrightarrow{\mathbf{h}}_s^{-1}\hat{\mathbf{g}}_s$, the vector of integers, $\hat{\mathbf{g}}_s = \{g_{a,s}, g_{b,s}, g_{c,s}\}$, indexes the small reciprocal space and $V_s = \det\overleftrightarrow{\mathbf{h}}_s$ is the volume of the small cell. The plane wave energy cutoff is taken to be $E_{cut}^{(\mathrm{short})}$ (with the cutoff on the density $4E_{cut}^{(\mathrm{short})}$).

Given that the electron density is localized in the small cell, the short range components of the Hartree and local pseudopotential energies can be evaluated straightforwardly,

$$\begin{aligned} E_{\mathrm{H}}^{(\mathrm{short})}[n] &= \frac{e^2}{2}\int_{D(\overleftrightarrow{\mathbf{h}}_s)} d\mathbf{r}\int_{D(\overleftrightarrow{\mathbf{h}}_s)} d\mathbf{r}'\,\frac{n_s(\mathbf{r})n_s(\mathbf{r}')\mathrm{erfc}(\alpha|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|} \\ &= \frac{e^2}{2V_s}{\sum_{\hat{\mathbf{g}}_s}}'\bar{n}_s(-\mathbf{g}_s)\bar{n}_s(\mathbf{g}_s)\left[\frac{4\pi}{g_s^2}\right]\left[1-\exp\left(-\frac{g_s^2}{4\alpha^2}\right)\right] + \frac{e^2\pi}{2V_s\alpha^2}|n_s(0)|^2 \end{aligned} \tag{12}$$

$$\begin{aligned} E_{\mathrm{loc}}^{(\mathrm{short})}[n] &= \sum_{J=1}^{N_s}\int_{D(\overleftrightarrow{\mathbf{h}}_s)} d\mathbf{r}\; n_s(\mathbf{r})\left[\phi_{\mathrm{loc},J}(\mathbf{r}-\mathbf{R}_J+\mathbf{R}_c) + \frac{eq_J\mathrm{erf}(\alpha|\mathbf{r}-\mathbf{R}_J+\mathbf{R}_c|)}{|\mathbf{r}-\mathbf{R}_J+\mathbf{R}_c|}\right] \\ &= \frac{1}{V_s}{\sum_{\hat{\mathbf{g}}_s}}'\sum_{J=1}^{N_s}\bar{n}_s^*(\mathbf{g}_s)\exp(-i\mathbf{g}_s\cdot[\mathbf{R}_J-\mathbf{R}_c]) \\ &\times\left[\tilde{\phi}_{\mathrm{loc},J}(\mathbf{g}_s) + \frac{4\pi eq_J}{g_s^2}\exp\left(-\frac{g_s^2}{4\alpha^2}\right)\right] \\ &+ \frac{1}{V_s}\sum_{J=1}^{N_s}\bar{n}_s(0)\left[\tilde{\phi}_{\mathrm{loc},J}^{(0)} - \frac{eq_J\pi}{\alpha^2}\right]. \end{aligned} \tag{13}$$

where the J sum runs over the N_s ions within the small cell, the $\hat{\mathbf{g}}_s$ sum runs over the large reciprocal-space grid of the small cell and $\mathbf{R}_c$ is the position of the small cell inside the large. Since the full system is not periodic on $\overleftrightarrow{\mathbf{h}}_s$ but on $\overleftrightarrow{\mathbf{h}}$, Eqs. (12-13) will only yield the correct short range energy if $\alpha V_s^{1/3} >> 1$ and $n(\mathbf{r}_s)$ vanishes on the small cell boundary. The non-local pseudopotential energy is short range and is assumed to be evaluated within the small cell (only, considering the N_s ions in the small cell and using the small cell reciprocal space). Similarly, the exchange correlation and the electronic kinetic energies can also be evaluated in the small cell using standard techniques.

Next, the expressions for the long range portions of the Hartree and local pseudopotential energies must be formulated. This can be accomplished by expanding the electron density localized in the small cell in terms of the plane waves of the large cell. This expansion is permitted because the electron density, localized in the small cell, obeys periodic boundary conditions in the large cell (i.e. it is zero on the surface of $\overleftrightarrow{\mathbf{h}}$). Thus,

$$\begin{aligned} E_{\rm H}^{\rm (long)}[n] &= \frac{e^2}{2}\sum_{\hat{\mathbf{S}}}\int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}\int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}'\ \frac{n(\mathbf{r})n(\mathbf{r}')\mathrm{erf}(\alpha|\mathbf{r}-\mathbf{r}'+\overleftrightarrow{\mathbf{h}}\hat{\mathbf{S}}|)}{|\mathbf{r}-\mathbf{r}'+\overleftrightarrow{\mathbf{h}}\hat{\mathbf{S}}|} \\ &= \frac{e^2}{2V}{\sum_{\hat{\mathbf{g}}}}'\bar{n}(-\mathbf{g})n(\mathbf{g})\left[\frac{4\pi}{g^2}\exp\left(-\frac{g^2}{4\alpha^2}\right)+\hat{\phi}^{(\rm screen,Coul)}(\mathbf{g})\right] \\ &+ \left(\frac{e^2}{2V}\right)\left[\hat{\phi}^{(\rm screen,Coul)}(0)-\frac{\pi}{\alpha^2}\right]|n(0)|^2 \end{aligned} \tag{14}$$

$$\begin{aligned} E_{\rm loc}^{\rm (long)}[n] &= -\sum_{\hat{\mathbf{S}}}\sum_{I=1}^{N}\int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}\ n(\mathbf{r})\left[\frac{eq_I\mathrm{erf}(\alpha|\mathbf{r}-\mathbf{R}_I+\overleftrightarrow{\mathbf{h}}\hat{\mathbf{S}}|)}{|\mathbf{r}-\mathbf{R}_I+\overleftrightarrow{\mathbf{h}}\hat{\mathbf{S}}|}\right] \\ &= -\frac{e}{V}{\sum_{\hat{\mathbf{g}}}}'\bar{n}^*(\mathbf{g})S(\mathbf{g})\left[\frac{4\pi}{g^2}\exp\left(-\frac{g^2}{4\alpha^2}\right)+\hat{\phi}^{(\rm screen,Coul)}(\mathbf{g})\right] \\ &- \frac{e}{V}\bar{n}_s(0)S(0)\left[\hat{\phi}^{(\rm screen,Coul)}(0)-\frac{\pi}{\alpha^2}\right]. \end{aligned} \tag{15}$$

where

$$S(\mathbf{g}) = \sum_I q_I\exp(i\mathbf{g}\cdot\mathbf{R}_I) \tag{16}$$

is the atomic structure factor and

$$\begin{aligned} \bar{n}(\mathbf{g}) &= \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}\exp[-i\mathbf{g}\cdot\mathbf{r}]n(\mathbf{r}) \\ &= \int_{D(\overleftrightarrow{\mathbf{h}_s})} d\mathbf{r}_s\exp[-i\mathbf{g}\cdot\mathbf{r}_s]n(\mathbf{r}_s+\mathbf{R}_c) \\ &= \int_{D(\overleftrightarrow{\mathbf{h}_s})} d\mathbf{r}_s\exp[-i\mathbf{g}\cdot(\mathbf{r}_s-\mathbf{R}_c)]n_s(\mathbf{r}_s) \end{aligned} \tag{17}$$

are the plane wave expansion coefficients of the electron density in the reciprocal space of the large cell, $\mathbf{g} = \overleftrightarrow{\mathbf{h}}^{-1}\ \hat{\mathbf{g}}$. The integral in Eq. (17) can be extended to cover the domain

described by the large cell without loss of generality because $n(\mathbf{r}_s + \mathbf{R}_c) \equiv 0$ outside of the small cell. Note, $\bar{n}(\mathbf{g}) = \bar{n}_s(\mathbf{g}_s)$ if $\overleftrightarrow{\mathbf{h}}_s \equiv \overleftrightarrow{\mathbf{h}}$ and $\mathbf{R}_c = 0$. Methods for the efficient evaluation of Eq. (17) and, hence, Eq. (14) and Eq. (15) are developed below.

First, it is clear from the long range/short range decomposition of the Hartree and local pseudopotential energies that a different plane wave cutoff can be introduced to treat each part. That is, one cutoff, $E_{cut}^{(\mathrm{short})}$, can be used to evaluate the short range components of the energy, Eq. (12) and Eq(13), and another, $E_{cut}^{(\mathrm{long})}$ can be used to evaluate the long range components, Eq. (14) and Eq.(15). While the long range/short range decomposition is general, it is expected that the short range contributions will be obtained by integration over functions that rapidly vary spatially while the long range contributions will be obtained by integration over a slowly varying function. Therefore, the short range energy contributions must be evaluated using a large reciprocal space cutoff (i.e. the standard $E_{cut}^{(\mathrm{density,short})} = 4E_{cut}^{(\mathrm{short})}$). In contrast, the long range part can be evaluated in reciprocal space using a small cutoff, $E_{cut}^{(\mathrm{long})} << E_{cut}^{(\mathrm{short})}$. Thus, by splitting the electronic energy into two parts, large gains in efficiency are possible.

Next, consider the case that the number of particles in the small cell, N_s and the small cell volume, V_s), are much less than their large cell counterparts ($N_s << N$ and $V_s << V$) as would be the case for a large, chemically inert bath surrounding a chemically active subsystem. The computational cost of evaluating the short range local pseudopotential and short range Hartree, exchange correlation, non-local pseudopotential and the electronic kinetic energy as well as the overlap matrix, $\langle\psi_{j,s}|\psi_{i,s}\rangle$, scales like $\sim N_s^3$. The computational cost of evaluating the long range part of the Hartree and local pseudopotential energies depends on the computational cost of evaluating the atomic charge density, S(**g**), and the plane wave expansion of the density in the large cell (see Eq. (17)). Since the atomic charge density can be evaluated in $N\log N$ using Particle Mesh Ewald techniques[15–17], if Eq. (17) could also be evaluated in $N\log N$, the computational cost of the method would then be $N\log N$ at fixed $\overleftrightarrow{\mathbf{h}}_s$ and N_s. (The present approach yields a linear scaling method because, at fixed particle density and plane wave cutoff, the number of plane waves increases linearly with particle number).

In order to achieve linear scaling, the electron density must be interpolated from the small cell where it is described by a plane wave expansion with a large cutoff, $E_{cut}^{(\mathrm{short})}$, to the large cell where it is described by a plane wave expansion with a small cutoff, $E_{cut}^{(\mathrm{long})}$, effectively. First, consider the Fourier components of the density

$$\bar{n}(\mathbf{g}) = \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \exp[-i\mathbf{g}\cdot\mathbf{r}] n(\mathbf{r}). \tag{18}$$

If $n(\mathbf{r})$ can be expressed in a finite plane wave basis,

$$n(\mathbf{r}) \equiv \frac{1}{V} \sum_{\hat{g}_a=-P_a/2+1}^{P_a/2} \sum_{\hat{g}_b=-P_b/2+1}^{P_b/2} \sum_{\hat{g}_c=-P_c/2+1}^{P_c/2} \exp(i\mathbf{g}\cdot\mathbf{r})\bar{n}(\mathbf{g}), \tag{19}$$

then the Fourier coefficients can also be determined (exactly) from a discrete sum over a

real space grid

$$\bar{n}(\mathbf{g}) \equiv \frac{V}{P_a P_b P_c} \sum_{\hat{s}_a=0}^{P_a-1} \sum_{\hat{s}_b}^{P_b-1} \sum_{\hat{s}_c=0}^{P_c-1} e^{-2\pi i \hat{g}_a \hat{s}_a/P_a} e^{-2\pi i \hat{g}_b \hat{s}_b/P_b} e^{-2\pi i \hat{g}_c \hat{s}_c/P_c} n(\overleftrightarrow{\mathbf{hs}}) \tag{20}$$

Here, P_a, P_b, and P_c are both the number of reciprocal lattice points along each direction and the number of points discretizing the $\mathbf{a}, \mathbf{b}, \mathbf{c}$ axes of the cell, and $s_\alpha = \hat{s}_\alpha / P_\alpha$. Importantly, Eq. (20) and its inverse, Eq. (19), can be evaluated using a three dimensional Fast Fourier Transforms (3D-FFT) in order $N \log N$. A spherical cutoff is introduced in reciprocal space by simply assuming that $n(\mathbf{r})$ is described by a basis in which $\bar{n}(\mathbf{g}) \equiv 0$ when $\hbar^2 |\mathbf{g}|^2 / 2m_e > E_{cut}$.

Next, consider a function, $f(\mathbf{r})$ with plane wave expansion coefficients,

$$\begin{aligned} \bar{f}(\mathbf{g}) &= \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \ \exp[-i\mathbf{g}\cdot\mathbf{r}] f(\mathbf{r}) \\ &= V \int_0^1 ds_a \int_0^1 ds_b \int_0^1 ds_c e^{-2\pi i \hat{g}_a s_a} e^{-2\pi i \hat{g}_b s_b} e^{-2\pi i \hat{g}_c s_c} f(\overleftrightarrow{\mathbf{hs}}). \end{aligned} \tag{21}$$

that can be described on a finite reciprocal space (cf. Eq. (20)). In order to express the plane wave expansion coefficients, accurately, in terms of a sum over an arbitrary set of equally spaced discrete points in real space (as opposed to the continuous integrals given in Eq. (21) or the discretization required by Eq. (20)), it useful to introduce the Euler exponential spline

$$\exp\left(\frac{2\pi i \hat{g}_\alpha u}{\tilde{P}_\alpha}\right) = d_m(\hat{g}_\alpha, \tilde{P}_\alpha) \sum_{\hat{s}=-\infty}^{\infty} M_m(u-\hat{s}) \exp\left(\frac{2\pi i \hat{g}_\alpha \hat{s}}{\tilde{P}_\alpha}\right) + \mathcal{O}\left(\frac{2|\hat{g}_\alpha|}{\tilde{P}_\alpha}\right)^m$$

$$d_m(\hat{g}_\alpha, \tilde{P}_\alpha) = \frac{\exp\left(2\pi i (m-1)/\tilde{P}_\alpha\right)}{\left[\sum_{j=0}^{m-2} M_m(j+1) \exp\left(2\pi i \hat{g}_\alpha j / \tilde{P}_\alpha\right)\right]} \tag{22}$$

where $\hat{s}$ is an integer, u is a real number, m is the spline order assumed to be even and the $M_m(u)$ are the Cardinal B splines

$$M_2(u) = 1 - |u-1| \tag{23}$$

$$\begin{aligned} M_m(u) &= \left[\frac{u}{m-1}\right] M_{m-1}(u) + \left[\frac{m-u}{m-1}\right] M_{m-1}(u-1) \\ M_m(u) &\neq 0 \qquad\qquad 0 < u < m \\ M_m(u) &= 0 \qquad\qquad u \leq 0, u \geq m \end{aligned} \tag{24}$$

The Cardinal B splines satisfy the following sum rule and recursion relation:

$$\sum_{\hat{s}=\infty}^{\infty} M_m(u-\hat{s}) = 1$$

$$\frac{dM_m(u)}{du} = M_{m-1}(u) - M_{m-1}(u-1)$$

Inserting the Euler exponential spline into Eq. (21) yields a well defined approximation to $\bar{f}(\mathbf{g})$,

$$\bar{f}(\mathbf{g}) \approx \left[V d_m^*(\hat{g}_a, \tilde{P}_a) d_m^*(\hat{g}_b, \tilde{P}_b) d_m^*(\hat{g}_c, \tilde{P}_c) \right] \times \sum_{\hat{s}_a=0}^{\tilde{P}_a-1} \sum_{\hat{s}_b=0}^{\tilde{P}_b-1} \sum_{\hat{s}_c=0}^{\tilde{P}_c-1} e^{-2\pi i \hat{g}_a \hat{s}_a/\tilde{P}_a} e^{-2\pi i \hat{g}_b \hat{s}_b/\tilde{P}_b} e^{-2\pi i \hat{g}_c \hat{s}_c/\tilde{P}_c} f^{(\mathrm{conv})}(\overleftrightarrow{\mathbf{h}}\mathbf{s}) \tag{25}$$

where

$$f^{(\mathrm{conv})}(\overleftrightarrow{\mathbf{h}}\mathbf{s}) = \int_0^1 ds'_a \int_0^1 ds'_b \int_0^1 ds'_c \sum_{k_a=-\infty}^{\infty} \sum_{k_b=-\infty}^{\infty} \sum_{k_c=-\infty}^{\infty} f(\overleftrightarrow{\mathbf{h}}\mathbf{s}') \times M_m([s'_a - k_a]\tilde{P}_a - \hat{s}_a) M_m([s'_b - k_b]\tilde{P}_b - \hat{s}_b) M_m([s'_c - k_c]\tilde{P}_c - \hat{s}_c). \tag{26}$$

is the interpolation of $f(\mathbf{r})$ onto the discrete real space grid defined by $s_\alpha = \hat{s}_\alpha/\tilde{P}_\alpha$ and $0 \le \hat{s}_\alpha \le \tilde{P}_\alpha - 1$.

Equation (25) can be evaluated using a 3D-FFT in order $N \log N$ provided the function, $f^{(\mathrm{conv})}(\overleftrightarrow{\mathbf{h}}\mathbf{s})$, defined on the discrete real space, can be constructed in a computationally efficient manner. In addition, Eq. (25) is smooth and possesses $m-2$ continuous derivatives. Note, if $\tilde{P}_a > m+1$ then each point in the continuous space, $\{s'_a, s'_b, s'_c\}$, is mapped to m^3 unique points on the discrete grid indexed by $\{\hat{s}_a, \hat{s}_b, \hat{s}_c\}$ due to the finite support of the $M_m(p)$ (see Eq. (23)). Also, it is important to choose $\tilde{P}_\alpha > P_\alpha$ to reduce the error inherent in the interpolation (see Eq. (22)).

It is now a simple matter to generate a computationally efficient and well defined approximation to the Fourier coefficients, $\bar{n}(\mathbf{g})$, of an electron density $n(\mathbf{r})$ that is assumed to be nonzero only in the small cell described by $\overleftrightarrow{\mathbf{h}}_s$. First, given that $\bar{n}_s(\mathbf{g}_s)$, defined in Eq. (11), exists on a finite reciprocal space, the identity given in Eq. (20) holds. Thus, the discrete form of the density can be inserted into Eq. (26) and the integrals performed using trapezoidal rule integration with loss of generality to yield the desired interpolation from the small cell to the large cell,

$$n^{(\mathrm{conv})}(\overleftrightarrow{\mathbf{h}}\mathbf{s}) = \left[\frac{V_s}{V}\right] \left[\frac{1}{P_{a,s} P_{c,s} P_{c,s}}\right] \sum_{\hat{s}'_a=0}^{P_{a,s}-1} \sum_{\hat{s}'_b=0}^{P_{b,s}-1} \sum_{\hat{s}'_c=0}^{P_{c,s}-1} \sum_{k_a=-\infty}^{\infty} \sum_{k_b=-\infty}^{\infty} \sum_{k_c=-\infty}^{\infty} n_s(\overleftrightarrow{\mathbf{h}}_s \mathbf{s}') \times M_m([s'_a + S_{a,s} - k_a]\tilde{P}_a - \hat{s}_a) M_m([s'_b + S_{b,s} - k_b]\tilde{P}_b - \hat{s}_b) \times M_m([s'_c + S_{c,s} - k_c]\tilde{P}_c - \hat{s}_c). \tag{27}$$

Here, $\{P_{a,s}, P_{b,s}, P_{c,s}\}$ are defined by the size of the small cell reciprocal space (through the plane wave energy cutoff, $E_{cut}^{(\mathrm{short})}$), $s'_\alpha = \hat{s}'_\alpha / P_{\alpha,s}$, $\mathbf{S}_s = \overleftrightarrow{\mathbf{h}}^{-1} \mathbf{R}_c$, and $V_s/V = \det \overleftrightarrow{\mathbf{D}}$ while the $\{\tilde{P}_a, \tilde{P}_b, \tilde{P}_c\}$ are defined by the size of the large cell reciprocal space (through the energy cutoff, $E_{cut}^{(\mathrm{long})}$).

The desired plane wave expansion of the density, $\bar{n}(\mathbf{g})$, is constructed by inserting $n^{(\mathrm{conv})}(\overleftrightarrow{\mathbf{h}}\mathbf{s})$ into Eq. (25) and performing a 3D-FFT. Note, in the limit, $\tilde{P}_a = P_{a,s}$, $\tilde{P}_b = P_{b,s}$, $\tilde{P}_c = P_{c,s}$ or $E_{cut}^{(\mathrm{short})} = E_{cut}^{(\mathrm{long})}$, and $\overleftrightarrow{\mathbf{h}} = \overleftrightarrow{\mathbf{h}}_s$, then $\bar{n}_s(\mathbf{g}_s) \equiv \bar{n}(\mathbf{g})$ because Eq. (20) is exact for a finite reciprocal space and the Euler exponential splines are

exact at the knots. Importantly, Eq. (27) can be evaluated in order $N_s m^3$ **and** the (dense) discrete real space grid spanning the small cell, $\overleftrightarrow{\mathbf{h}}_s$, and the (sparse) discrete real space grid spanning the large cell, $\overleftrightarrow{\mathbf{h}}$, need not be commensurate. In addition, the separable form of the $M_m(p)$, which is a consequence of the choice $\overleftrightarrow{\mathbf{h}}^{-1}\overleftrightarrow{\mathbf{h}}_s = \overleftrightarrow{\mathbf{D}}$, allows the required $M_m(p)$ to be evaluated independently in order $mN_s^{1/3}$. Thus, the overall computational cost of constructing $\bar{n}(\mathbf{g})$ is $N\log N$ (dominated by the FFT). Finally, the resulting $\bar{n}(\mathbf{g})$ (i.e. obtained by inserting Eq. (27) into Eq. (25)) is continuously differentiable with respect to the expansion coefficients of the orbitals, $\bar{\psi}_{j,s}(\mathbf{g}_s)$, defined in Eq. (11).

1.2 Use in cluster calculations

In any *ab initio* molecular dynamics scheme in which long-range forces are computed in reciprocal space, e.g. plane-waves, Gaussians, DVRs, systems with reduced periodicity, e.g., clusters, surfaces, wires, can be treated using the screening function methodology developed by Martyna and Tuckerman[11–13]. Typically, when using the screening function, however, the size of the box needs to be roughly twice the maximum distance between the two furthest atoms in the cluster, which makes the use of this methodology somewhat more expensive than other techniques. The dual-gridding technique above can be used to circumvent this problem. In order to use the dual-gridding scheme in the context of a cluster calculation, for example, one simply uses two boxes (see Fig. 1): The central box contains the cluster system and is chosen large enough to contain the cluster with a small buffer region. The outer box can be chosen quite large, at least twice as large as the furthest distance between two atoms in the cluster (but it can be larger as well). The coarse grid is then used to describe the **g**-space density in the large box, and the scheme outlined in the Methods section is used to compute the long-range energies. The Hartree energy, for example, would be computed as

$$E_{\rm H} = \frac{1}{2V_s}\sum_{\mathbf{g}_s} |\bar{n}(\mathbf{g}_s)|^2 \tilde{\phi}^{(\rm short)}(\mathbf{g}_s) + \frac{1}{2V}\sum_{\mathbf{g}\neq(\mathbf{0},\mathbf{0},\mathbf{0})} |\bar{n}(\mathbf{g})|^2 \bar{\phi}^{(\rm long)}(\mathbf{g}) \qquad (28)$$

1.3 Illustrative examples

As a first, simple illustrative example, consider a simple Gaussian electron density,

$$n(\mathbf{r}) = \left(\frac{\kappa^2}{\pi}\right)^{3/2} \exp(-\kappa^2 r^2). \qquad (29)$$

The interaction of this density with a point charge located an arbitrary distance, r_0, away from its center can be determined, analytically, $E_{ext} = \mathrm{erf}(\kappa r_0)/r_0$. In Table I, the convergence of the total external energy to the analytical value is presented as a function of the large cell plane wave cutoff and Cardinal B-spline interpolation order, for various choices of r_0. The calculations were performed using the cluster boundary condition technique of reference[11] and fixed small cell plane wave cutoff ($E_{cut}^{(\rm short)} = 120$ Ry). In general, it can be seen that low B-spline interpolation orders and small plane wave cutoffs in the large

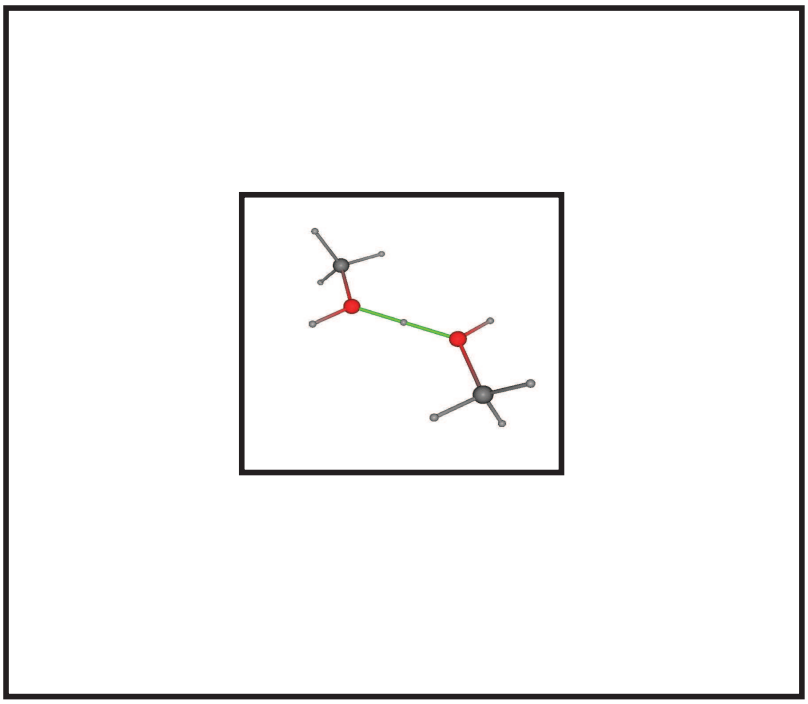


Figure 1. Illustration of the dual gridding scheme for cluster calculations.

cell, ($E_{cut}^{(\mathrm{long})}$), are sufficient to produce accurate results, indicating that the new method is both valid and efficient.

Next, we consider the human carbonic anhydrase II (HCA II) enzyme solvated in liquid water. In detail, the 260-residue HCA-II enzyme (complete with catalytic zinc – see Fig. 2), was solvated by 8,859 waters, for a total of 30,649 atoms. Clearly, a full *ab initio* treatment of such a large system is not feasible, at present. However, a hybrid model, wherein only the catalytic zinc, the side-chains of active site residues, HIS 94, HIS 96, HIS 119, THR 199, GLU 106 and the five water molecules in the active site are treated using an *ab initio* description, can be studied. Thus, 320 valence electrons of 80 atoms in the active site (see Fig. 2) are treated at an *ab initio* level while the remainder of the system is treated using the empirical CHARMM22 all-atom parameter force field which includes TIP3P water model[23]. Briefly, the electrons are assumed to interact with "*ab initio*" atoms via standard Troullier-Martins pseudopotentials[24] and with "empirical atoms" via pseudopotentials fit by the authors (see also[5,25]). The BLYP, density functional[20,21] was employed to treat exchange and correlation. *Ab initio* atoms (ion-cores) were permitted to interact with neighboring "empirical atoms" via appropriate bond, bend, torsion, one-four, van der Waals and Coulomb forces. The parameters were obtained by enforcing good agreement between mixed models, fully empirical models and fully *ab initio* models of relevant fragments. For example, the minimum energy geometry of hybrid model $\mathrm{CH_3CO-(HIS)-NHCH_3}$ deviates at most 2 degrees in the bend angles and 0.02Å in

Table 1. The interaction of a Gaussian charge density, $\kappa = 3.779454\mathring{A}^{-1}$, with a point charge at distance, r_0 from its center is presented as a function of large cell plane wave cutoff and B-Spline interpolation order. The large cell size was fixed at $L_l = 20\mathring{A}$ on edge. The small cell size was fixed at $L_s = 4\mathring{A}$ on edge and the small cutoff was fixed at $E_{cut}^{(\mathrm{short})} = 120Ry$. The electrostatic division parameter was set to be $\alpha = 6/L_s$ and $\Delta E_{ext} = E_{ext} - E_{ext}^{(exact)}$.

r_0 (Å)	$E_{cut}^{(\mathrm{long})}$ (Rydberg)	m	E_{ext} (Hartree)	ΔE_{ext} (Kelvin)
4	4	4	-0.132296	1
		6	-0.132297	1
		8	-0.132297	1
	8	4	-0.132293	0
		6	-0.132293	0
		8	-0.132293	0
6	4	4	-0.088186	3
		6	-0.088185	3
		8	-0.088185	3
	8	4	-0.088198	1
		6	-0.088198	1
		8	-0.088198	1
8	4	4	-0.066126	7
		6	-0.066125	7
		8	-0.066125	7
	8	4	-0.066149	1
		6	-0.066149	1
		8	-0.066149	1

the bond lengths from the standards (CHARMM and fully *ab initio* treatments as appropriate).

The HCA II/water system described above was prepared by taking the crystallographic configuration of the enzyme (PDB identification label, "1RAY")[26] and immersing it in TIP3P water. Next, a 1 ns constant temperature molecular dynamics calculation was performed using a fully empirical treatment[23]. This was followed by a 1 ns constant pressure molecule dynamics calculation. At this point, the hybrid model was introduced. In Table

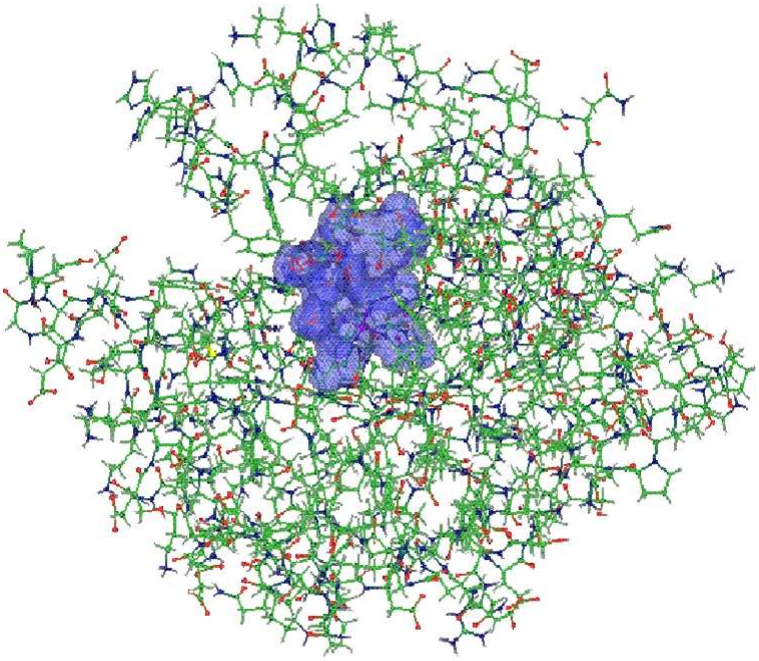


Figure 2. Snapshot of human carbonic anhydrase. The electronic density in the active site is shown as the blue isosurface.

Table 2. The total electronic energy of the active site of HCA II immersed in a bath of TIP3P molecules and CHARMM22 model amino acid residues as a function of large cell plane cutoff and spline interpolation order. The large cell size is fixed by the state point, 66.7Å, on edge. The small cell size was fixed at 18Å on edge and the small cell cutoff was fixed at 70Ry. The electrostatic division parameter was set to be $\alpha = 9/L_s$ and the accuracy measure is defined to be $\Delta E_{tot} = E_{tot}(E_{cut}^{(\mathrm{long})}, m) - E_{tot}(4, 8)$.

$E_{cut}^{(\mathrm{long})}$ (Rydberg)	m	E_{tot} (Hartree)	ΔE_{tot} (Kelvin)
0.5	6	-2329.31984	9200
	8	-2329.33018	5900
2	6	-2329.34896	32
	8	-2329.34905	3
4	6	-2329.34905	3
	8	-2329.34906	0

2, the convergence of the electronic energy for a representative configuration taken from the simulation of the hybrid model, is shown versus the large cell, plane wave cutoff and the Cardinal B-spline interpolation order. As is clear from the table, accurate energies are obtained for low spline orders and plane wave cutoffs.

2 Exploration of Free-Energy Surfaces

One of the key quantities in thermodynamics is the free energy associated with changes in conformation or thermodynamic state of a complex system. Molecular dynamics (MD) and Monte Carlo based approaches have emerged as important theoretical tools to study such free-energy changes along one-dimensional paths, e.g. along single reaction coordinates or one-dimensional λ-switching paths. Among these approaches, Umbrella Sampling[27–29] and Thermodynamic Integration[30–32] remain the most popular because they can be easily implemented. However, the problem of computing a multi-dimensional free-energy surface (FES) in several collective variables or reaction coordinates of interest has remained a significant challenge, particular when the FES contains numerous minima separated by high barriers. The mapping out of the free-energy landscape of small peptides and proteins in the Ramachandran angles, radius of gyration and/or number of hydrogen bonds, or the characterization of dissociation or mass-transfer processes in aqueous solution in terms of coordination numbers and distances are examples of this type of problem.

The challenge of treating such "rough" energy landscapes has lead to the introduction of various important new techniques for enhanced sampling of the configurational distribution of complex systems, from which the free energy is obtained. These include parallel tempering[33–38], hyperdynamics[39], parallel replica dynamics[40], Wang-Landau sampling[41,42], configuration-bias 'Monte Carlo[43], the Reference-Potential Spatial-Warping Algorithm[44,45], metadynamics[46], and techniques based on adiabatic dynamics[47–51], as a few examples. A comprehensive review of free-energy techniques was recently presented in the edited volume, *Free Energy Calculations*[52].

The Adiabatic Free Energy Dynamics (AFED),[47–49] introduced eight years ago by Rosso, *et al.*,[53] is a dynamical scheme for generating free-energy hypersurfaces in several collective variables of interest. The approach employs an imposed adiabatic decoupling between a small set of collective variables or reaction coordinates and the remaining degrees of freedom. Within this scheme, an elevated temperature is also applied to the collective variables to ensure that they are able to cross the high energy barriers needed to ensure sufficient sampling. In the limit of high temperature and adiabatic decoupling, it can be shown that the free energy hypersurface in the collective variables is obtained directly from their resultant probability distribution function[47,48]. The approach has been applied to the conformational sampling of small peptides[49] and in the computation of solvation and binding free energies via alchemical transformations[51]. In both cases, the use of adiabatic dynamics has been shown to lead to significant improvement in efficiency compared to traditional methods such as free-energy perturbation[54], umbrella sampling[27–29], and the blue moon ensemble approach[31,32]. In addition to being a relatively fast method, the AFED approach requires no *a posteriori* processing of the simulation data. Moreover, AFED is able to generate multi-dimensional free-energy hypersurfaces with significantly greater efficiency than multidimensional versions of the aforementioned approaches[49]. By construction, AFED generates full sweeps of the free-energy surface and, therefore, can

rapidly map out the locations of the free-energy minima well before the entire surface is fully converged.

The AFED approach is derived and implemented, in practice, by transforming the coordinate integrations in the canonical partition function to a set of generalized coordinates that explicitly contain the collective variables of interest. This gives rise to the disadvantage that the adiabatic dynamics must be carried out in these generalized coordinates, which leads to a steep implementation curve due to the rather invasive modifications to existing MD packages needed to introduce these transformations. It should be noted, however, that once such transformations are put in place, they can be subsequently combined with additional spatial-warping transformations that also significantly enhance conformational sampling[44,45].

Recently, Maragliano and Vanden-Eijnden[50] and, independently, Abrams and Tuckerman[55] built on the AFED approach by introducing a set of extended phase-space or "driving" variables that are harmonically coupled to the collective variables of interest. By imposing the adiabatic decoupling and high temperature on these extended variables rather than on the collective variables, the need for explicit transformations is avoided, thereby enlarging the class of collective variables that can be treated and rendering the technique substantially easier to implement. Maragliano and Vanden-Eijnden named the new technique "temperature accelerated molecular dynamics" or TAMD while Abrams and Tuckerman named if driven-AFED or d-AFED. It should be noted that such "driving" variables are also central in the so-called "metadynamics" approach[46], where they are used together with a time-dependent potential that floods energy basins with Gaussians, thereby allowing the system to escape the basin and move into a neighboring one. In metadynamics, as the basins are filled, the histogram in the collective variables becomes flat. When this occurs, the sum of all of the Gaussians is used to recover the free-energy hypersurface.

2.1 Adiabatic free-energy dynamics

Consider a system of N particles with Cartesian coordinates $\mathbf{r}_1, ..., \mathbf{r}_N \equiv \mathbf{r}$ and conjugate momenta $\mathbf{p}_1, ..., \mathbf{p}_N \equiv \mathbf{p}$ subject to a potential energy $V(\mathbf{r}_1, ..., \mathbf{r}_N)$. The classical canonical partition function for the system is given by

$$Q = C \int d^N\mathbf{p} \int_{D(V)} d^N\mathbf{r} \exp\left\{-\beta\left[\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + V\left(\mathbf{r}_1, \ldots, \mathbf{r}_N\right)\right]\right\} \tag{30}$$

where $H(\mathbf{p}, \mathbf{r}) = \sum_i \mathbf{p}_i^2/2m_i + V(\mathbf{r})$ is the Hamiltonian, $\beta = 1/k_BT$, $D(V)$ is the spatial domain defined by the containing volume, and C is an overall prefactor that renders Q dimensionless and compensates for overcounting of states obtained by exchanging particles of the same chemical identity. For a system with M species, $C = [h^{3N} \prod_{\alpha=1}^{M} N_\alpha!]^{-1}$, where h is Planck's constant, and N_α is the number of particles of species α.

Suppose we wish to determine the free-energy hypersurface in a set of $n < N$ collective variables $q_1(\mathbf{r}), ..., q_n(\mathbf{r})$. Examples are the Ramachandran angles for characterizing the conformational space of oligopeptides or combinations of distances for characterizing a chemical reaction. The probability that $q_1(\mathbf{r})$ has the value s_1, $q_2(\mathbf{r})$ has the value s_2,...,$q_n(\mathbf{r})$ has the value s_n is given by

$$P(s_1, ..., s_n) = \frac{\int d^N\mathbf{p} d^N\mathbf{r}\ e^{-\beta H(\mathbf{p},\mathbf{r})} \prod_{i=1}^{n} \delta(q_i(\mathbf{r}) - s_i)}{\int d^N\mathbf{p} d^N\mathbf{r}\ e^{-\beta H(\mathbf{p},\mathbf{r})}} \tag{31}$$

Given this probability distribution, the free-energy hypersurface can be calculated according to

$$F(s_1, ..., s_n) = -kT \ln P(s_1, ..., s_n) \tag{32}$$

In many complex systems, direct calculation of the probability distribution function from a molecular dynamics trajectory is intractable because of the existence of high free-energy barriers separating important minima on the hypersurface. Free-energy surfaces of this type are said to be "rough", and it is necessary to employ enhanced sampling techniques. The adiabatic free-energy dynamics (AFED) achieves enhanced sampling in the variables $q_1(\mathbf{r}), ..., q_n(\mathbf{r})$ by introducing a high temperature $T_s \gg T$ for these n degrees of freedom only, while maintaining the remaining $3N - n$ degrees of freedom at the correct ensemble temperature T. The temperature disparity can be accomplished by introducing two separate sets of thermostats for each set of degrees of freedom. The high temperature T_s ensures that the variables $q_1, ..., q_n$ are able to cross high energy barriers on their part of the energy landscape. However, this high temperature also destroys the thermodynamic properties of the system *unless* the variables $q_1, ..., q_n$ are also adiabatically decoupled from the remaining degrees of freedom. In order to accomplish this decoupling, we need to be able to run the dynamics in a coordinate system that explicitly contains $q_1, ..., q_n$.

Suppose there is a transformation from Cartesian coordinates $\mathbf{r}_1, ..., \mathbf{r}_N$ to generalized coordinates $q_1, ..., q_{3N} \equiv q$ via the transformation equations $q_\alpha = q_\alpha(\mathbf{r})$. The inverse transformations are denoted $\mathbf{r}_i = \mathbf{r}_i(q)$. Substituting the transformation into Eq. (30) yields

$$\begin{aligned} Q &= C \int d^N \mathbf{p} \int_{D(V)} d^{3N} q \; J(q) \exp \left\{ -\beta \left[\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + V\left(\mathbf{r}_1(q), ..., \mathbf{r}_N(q)\right) \right] \right\} \\ &= C \int d^N \mathbf{p} \int_{D(V)} d^{3N} q \exp \left\{ -\beta \left[\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + \tilde{V}\left(q_1, ..., q_{3N}\right) \right] \right\} \end{aligned} \tag{33}$$

where the potential $\tilde{V}$ contains the Jacobian of the transformation $J(q) = |\partial \mathbf{r} / \partial q|$ and is given by

$$\tilde{V}(q_1, ..., q_{3N}) = V(\mathbf{r}_1(q), ..., \mathbf{r}_N(q)) - kT \ln J(q_1, .., q_{3N}) \tag{34}$$

Note that the partition function in Eq. (33) is completely equivalent to that in Eq. (30). Moreover, even though the transformation is not canonical, since we are only interested in sampling the ensemble distribution, we can treat the $3N$ Cartesian momentum components as "conjugate" to the $3N$ generalized coordinates, each being defined as $p_\alpha = m_\alpha \dot{q}_\alpha$, $\alpha = 1, ..., 3N$, where m_α are the associated masses. Thus, in order to achieve the desired adiabatic decoupling, we simply choose the first n masses m_α to be much larger than all of the remaining masses, $m_{1,...,n} \gg m_{n+1,...,3N}$.

Under the conditions of adiabatic decoupling and the temperature disparity, it was shown in Refs.[47,48], via a decomposition of the classical propagator, that the probability distribution, denoted $P_{\mathrm{adb}}(s_1, ..., s_n)$, becomes

$$P_{\mathrm{adb}}(s_1, ..., s_n) = \mathcal{N} \int d^n p \exp \left[-\beta_s \sum_{\alpha=1}^{n} \frac{p_\alpha^2}{2m_\alpha} \right] \left[Z(s_1, ..., s_n, \beta) \right]^{T/T_s} \tag{35}$$

where

$$Z(s_1,...,s_n,\beta) = \int d^{3N-n}p d^{3N}q \ \exp\left\{-\beta\left[\sum_{\alpha=n+1}^{3N}\frac{p_\alpha^2}{2m_\alpha} + \tilde{V}(q_1,...,q_{3N})\right]\right\} \times \prod_{\alpha=1}^{n}\delta(q_\alpha - s_\alpha) \tag{36}$$

and $\mathcal{N}$ is an overall normalization factor. In this case, because of the temperature ratio T/T_s in the exponent, the exact free-energy $F(s_1,...,s_n)$ at the temperature T, which is defined to be $F(s_1,...,s_n) = -kT\ln Z(s_1,...,s_n,\beta)$, is obtained from $P_{\rm adb}(s_1,...,s_n)$ by

$$F(s_1,...,s_n) = -kT_s \ln P_{\rm adb}(s_1,...,s_n) \tag{37}$$

Note that the multiplicative factor $-kT_s$ in Eq. (37) ensures that the free energy *at temperature* T is obtained. Eq. (37) shows that the free-energy surface can be computed *directly* from the probability distribution function generated in an adiabatic dynamics calculation. A detailed proof of the AFED method is given in Refs.[47,48].

2.2 Adiabatic free-energy dynamics without transformations

The AFED approach is a powerful one that is capable of generating multidimensional free-energy surfaces efficiently, as was shown in Refs.[48,49]. However, the need to work in generalized coordinates is a distinct disadvantage of the method, as this requires rather invasive modifications to existing molecular dynamics codes.

Recently, Maragliano and Vanden-Eijnden[50] and Abrams and Tuckerman[55] showed that AFED could be re-expressed in a set of extended phase-space variables in a manner similar to that used in the metadynamics approach of Laio and Parrinello[46], thereby circumventing the need for explicit coordinate transformations. This new formulation, which the authors called "Temperature Accelerated Molecular Dynamics" (TAMD) or "driven-AFED" (d-AFED) increases both the flexibility of the AFED method, allowing larger classes of collective variables to be treated, and the ease of implementation in existing packages.

TAMD/d-AFED can be derived as follows. We rewrite the product of δ-functions in Eq. (31) as the limit of a product of Gaussian functions[56]

$$\prod_{\alpha=1}^{n}\delta\left(q_\alpha(\mathbf{r}) - s_\alpha\right) = \lim_{\kappa\to\infty}\sqrt{\frac{\beta\kappa}{2\pi}}\exp\left[-\sum_{\alpha=1}^{n}\frac{\beta}{2}k\left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2\right] \tag{38}$$

When Eq. (38) is substituted into Eq. (31), we obtain

$$P(s_1,...,s_n) = \lim_{\kappa\to\infty}\mathcal{N}_\kappa\int d^N\mathbf{p}\int_{D(V)} d^N\mathbf{r} \times \exp\left\{-\beta\left[H(\mathbf{p},\mathbf{r}) + \frac{1}{2}\kappa\sum_{\alpha=1}^{n}\left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2\right]\right\} \tag{39}$$

where $\mathcal{N}_\kappa$ is a κ-dependent normalization constant. For large but finite κ, the integral in Eq. (39) represents a close approximation to the true probability distribution, and we

can regard the harmonic term in Eq. (39) as an additional potential term that keeps the collective variables $q_1(\mathbf{r}), ..., q_n(\mathbf{r})$ close to the values $s_1, ..., s_n$. In this representation, Eq. (39) resembles the probability distribution generated within the umbrella sampling approach[27–29]. However, if a set of n independent Gaussian integrations is introduced into Eq. (39) in the following form

$$P(s_1, ..., s_n) = \lim_{\kappa \to \infty} \mathcal{N}'_\kappa \int d^N\mathbf{p} \int_{D(V)} d^N\mathbf{r} \times \exp\left\{-\beta\left[H(\mathbf{p}, \mathbf{r}) + \sum_{\alpha=1}^{n} \frac{p_{s_\alpha}^2}{2m_\alpha} + \frac{1}{2}\kappa \sum_{\alpha=1}^{n} \left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2\right]\right\} \quad (40)$$

then the dependence of the distribution on $s_1, ..., s_n$ remains unaltered.

The argument of the exponential can now be regarded as an extended phase-space Hamiltonian

$$H_{\mathrm{ex}}(\mathbf{p}, p_s, \mathbf{r}, s) = \sum_{\alpha=1}^{n} \frac{p_{s_\alpha}^2}{2m_\alpha} + \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{r}_1, ..., \mathbf{r}_N) + \sum_{\alpha=1}^{n} \frac{1}{2}\kappa \left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2 \quad (41)$$

This Hamiltonian generates the dynamics of the original N Cartesian positions and momenta and of the additional n variables $s_1, ..., s_n \equiv s$ and their conjugate momenta $p_{s_1}, ..., p_{s_n} \equiv p_s$ The extended variables serve to "drag" or "drive" the collective variables $q_1(\mathbf{r}), ..., q_n(\mathbf{r})$ via the harmonic coupling through their portion of the energy landscape provided that the variables $s_1, ..., s_n$ are able to sample a comparable region.

Assuming, again, that there are significant barriers hindering the sampling of the collective variables, enhanced sampling can be achieved by employing a high temperature and adiabatic decoupling, this time on the extended phase-space variables[50]. Thus, we introduce a temperature $T_s \gg T$ and masses $m_\alpha \gg m_i$ for these variables. As in the original AFED scheme, the former condition ensures that high barriers can be crossed, if T_s is chosen high enough, while the large masses ensure adiabatic decoupling of the extended phase-space variables from all other degrees of freedom. Following Refs.[47,48,51], it can be shown that, under these conditions, the distribution function generated takes the form

$$P_{\mathrm{adb}}^{(\kappa)}(s_1, ..., s_n) \propto \int d^n p \exp\left[-\beta_s \sum_{\alpha=1}^{n} \frac{p_{s_\alpha}^2}{2m_\alpha}\right] \left[Z(s_1, ..., s_n, \beta)\right]^{\beta_s/\beta} \quad (42)$$

where $\beta_s = 1/kT_s$ and

$$Z(s_1, .., s_n, \beta) = \int d^N\mathbf{p} \int_{D(V)} d^N\mathbf{r} \, \exp\left\{-\beta\left[\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + \bar{V}(\mathbf{r}, s)\right]\right\} \quad (43)$$

and

$$\bar{V}(\mathbf{r}, s) = V(\mathbf{r}_1, ..., \mathbf{r}_N) + \frac{1}{2}\kappa \sum_{\alpha=1}^{n} \left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2 \quad (44)$$

The probability distribution in Eq. (42) generates an approximation $F_\kappa(s_1, ..., s_n)$ to the true free energy profile at temperature T according to

$$F_\kappa(s_1, ..., s_n) = -kT_s \ln P_{\mathrm{adb}}^{(\kappa)}(s_1, ..., s_n) \quad (45)$$

and it is clear that in the limit $\kappa \to \infty$, the true free energy profile is recovered

$$F(s_1, ..., s_n) = \lim_{\kappa \to \infty} F_\kappa(s_1, ..., s_n) \tag{46}$$

Eqs. (45) and (46) show that the free-energy hypersurface can be generated within the adiabatic dynamics scheme without requiring a transformation to generalized coordinates.

The ability of TAMD and d-AFED to generate the free-energy surface efficiently depends on the thermostatting mechanism employed to maintain the two temperatures. The adiabatic decoupling represents a non-equilibrium steady state, and in Refs.[47,48], it was shown that the generalized Gaussian moment thermostat (GGMT) of Liu and Tuckerman[57] is an effective approach for maintaining the temperature disparity within the AFED scheme. Therefore, we employ it here as well. For completeness, we show the explicit equations of motion, including the coupling to separate GGMTs at temperatures T and T_s. As noted, GGMTs are capable of maintaining temperature control under the nonequilibrium (steady-state) conditions implied by the two temperatures and adiabatic decoupling. Within the two-moment version of the GGMT technique, with a separate thermostat coupled to each degree of freedom, the equations of motion for the d-AFED scheme read

$$\begin{aligned}
\dot{r}_{i,k} &= \frac{p_{i,k}}{m_i} \\
\dot{p}_{i,k} &= F_{i,k} - \kappa \sum_{\alpha=1}^{n} \left(q_\alpha(\mathbf{r}) - s_\alpha\right) \frac{\partial q_\alpha}{\partial r_{i,k}} - \frac{p_{\eta_{i,k,1}}}{Q_1} p_{i,k} - \frac{p_{\eta_{i,k,2}}}{Q_2} \left[(kT) p_{i,k} + \frac{p_{i,k}^3}{3m_i}\right] \\
\dot{s}_\alpha &= \frac{p_{s_\alpha}}{m_\alpha} \\
\dot{p}_{s_\alpha} &= \kappa \left(q_\alpha(\mathbf{r}) - s_\alpha\right) - \frac{p_{\xi_{\alpha,1}}}{Q_1'} p_{s_\alpha} - \frac{p_{\xi_{\alpha,2}}}{Q_2'} \left[(kT_s) p_{s_\alpha} + \frac{p_{s_\alpha}^3}{3m_\alpha}\right] \\
\dot{\eta}_{i,k,1} &= \frac{p_{\eta_{i,k,1}}}{Q_1} \\
\dot{\eta}_{i,k,2} &= \left[(kT) + \frac{p_{i,k,1}^2}{m_i}\right] \frac{p_{\eta_{i,k,2}}}{Q_2} \\
\dot{\xi}_{\alpha,1} &= \frac{p_{\xi_{\alpha,1}}}{Q_1'} \\
\dot{\xi}_{\alpha,2} &= \left[(kT_s) + \frac{p_{s_\alpha}^2}{m_\alpha}\right] \frac{p_{\xi_{\alpha,2}}}{Q_2'} \\
\dot{p}_{\eta_{i,k,1}} &= \frac{p_{i,k}^2}{m_i} - kT \\
\dot{p}_{\eta_{i,k,2}} &= \frac{p_{i,k}^4}{3m_i^2} - (kT)^2 \\
\dot{p}_{\xi_{\alpha,1}} &= \frac{p_{s_\alpha}^2}{m_\alpha} - kT_s \\
\dot{p}_{\xi_{\alpha,2}} &= \frac{p_{s_\alpha}^4}{3m_\alpha^2} - (kT_s)^2
\end{aligned} \tag{47}$$

where $F_{i,k} = -\partial V / \partial r_{i,k}$. In Eqs. (47), the thermostats are used to control the fluctuations in the second and fourth moments of the distribution of each momentum variable in the sys-

tem, whether these correspond to T or T_s. Here, k indexes the three Cartesian components of the physical coordinate and momentum of particle i, and $\eta_{i,k,1}$, $\eta_{i,k,2}$, $p_{\eta_{i,k,1}}$ and $p_{\eta_{i,k,2}}$ are the corresponding GGMT variables. Similarly, α indexes the extended phase-space driving variables, and $\xi_{\alpha,1}$, $\xi_{\alpha,2}$, $p_{\xi_{\alpha,1}}$ and $p_{\xi_{\alpha,2}}$ are the corresponding GGMT variables. The thermostat mass parameters Q_1, Q_2, Q_1', and Q_2' are chosen according to[57]:

$$\begin{aligned} Q_1 &= kT\tau^2 \qquad & Q_2 &= \frac{8}{3}(kT)^3\tau^2 \\ Q_1' &= kT_s\tau_s^2 \qquad & Q_2' &= \frac{8}{3}(kT_s)^3\tau_s^2 \end{aligned} \tag{48}$$

where τ and τ_s are characteristic time scales in the physical and extended systems, respectively. A typical choice for τ_s is the period of the harmonic coupling $(2\pi)\sqrt{m_\alpha/\kappa}$.

Another important feature of Eqs. (47) is that the presence of the stiff harmonic force term $(\kappa/2)\sum_\alpha (q_\alpha(\mathbf{r}) - s_\alpha)^2$ renders them amenable to multiple time-scale (r-RESPA) integration techniques[58,59]. For Eqs. (47), the Liouville operator, from which such integrators are derived, can be subdivided into a reference system containing the stiff oscillations of the harmonic coupling and a second propagator for the relatively slow motions associated with the motion of the physical system. Using r-RESPA to evaluate the inexpensive, but very fast, stiff harmonic force with a smaller timestep, while keeping the fundamental timestep larger, significantly improves the efficiency of the method. For details of this type of factorization the interested reader is directed to reference[58,59].

2.3 An illustrative example

The alanine hexamer (N-acetyl-(alanine)$_6$-methylamide) is a six-residue peptide that exhibits helical properties in solution[60]. Furthermore, computational studies of the alanine hexamer in solution, parameterized with the AMBER force field, results in helical conformational minima[61].

The simulation was performed by solvating the alanine hexamer in 698 TIP3P water molecules in a 27.9737 Å cubic periodic box. The molecule was started in its completely extended conformation and then equilibrated for 200 ps at constant volume (NVT), 1 ns at constant pressure (NPT), and finally 500 ps at constant volume (NVT).

In applying the d-AFED/TAMD method with the AMBER (parm94) force field, the two collective variables of interest chosen were the radius of gyration (R_G) and the number of intramolecular hydrogen bonds (N_H). These collective variables are defined as follows[62]

$$\begin{aligned} R_G &= \sqrt{\frac{1}{N_b}\sum_{i=1}^{N_b}\left(\mathbf{r}_i - \frac{1}{N_b}\sum_{j=1}^{N_b}\mathbf{r}_j\right)^2} \\ N_H &= \sum_{i=1}^{N_{\mathrm{Ox}}}\sum_{j=1}^{N_{\mathrm{Hy}}}\frac{1-\left(\frac{\mathbf{r}_i-\mathbf{r}_j}{d_0}\right)^6}{1-\left(\frac{\mathbf{r}_i-\mathbf{r}_j}{d_0}\right)^{12}} \end{aligned} \tag{49}$$

where N_b is the number of heavy backbone atoms, where N_{Ox} and N_{Hy} are the number of oxygen and hydrogen atoms, respectively, and $d_0 = 2.5$Å. Corresponding extended coordinates, s_1 and s_2, were added and these coordinates were treated as the slow variables

with masses $m_{s_1,s_2} = 15m_C$, where m_C is the mass of a carbon atom, and heated to a temperature of $T_{s_1,s_2} = 600$ K, while the physical variables were kept at a temperature of 300 K. s_1 and s_2 were coupled to R_G and N_H with harmonic coupling constants 5.4×10^6 K $\cdot$ Å^{-2} and 5.4×10^6 K, respectively.

The free energy surface computed from a relatively short production run of 5 ns, for the alanine hexamer in solution, leads to the free energy surface shown in Figure 3. The most significant feature of this surface is the presence of a global minimum at $(R_G, N_H) = (3.8, 4.4)$, corresponding to a right-handed helical, α_{R} , conformation (Figure 4a). Furthermore, there is also the presence of a large region of deformed/partial helices all with free energies lower than $+2$ kcal $\cdot$ mol^{-1} above the global minimum. These preliminary results are very consistent with previous results[61] for the alanine hexamer using the AMBER force field.

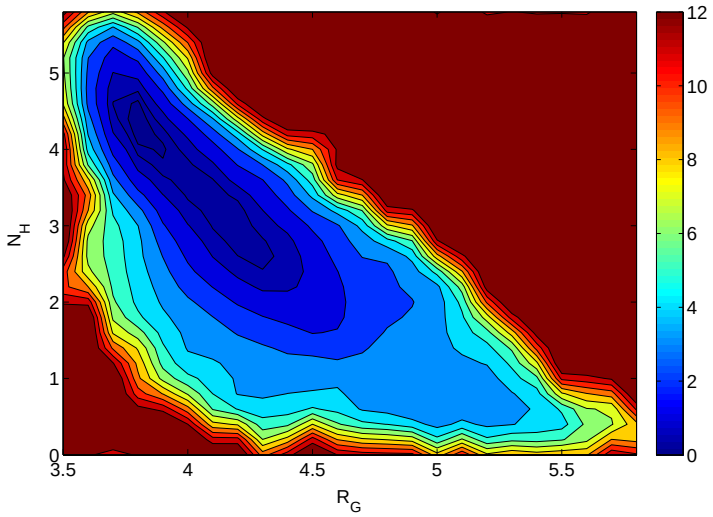


Figure 3. (a) Free-energy surface $F(R_G,\ N_H)$ and (b) contour plot computed for N-acetyl-(alanine)$_6$-methylamide (alanine hexamer) in solution with the AMBER force field.

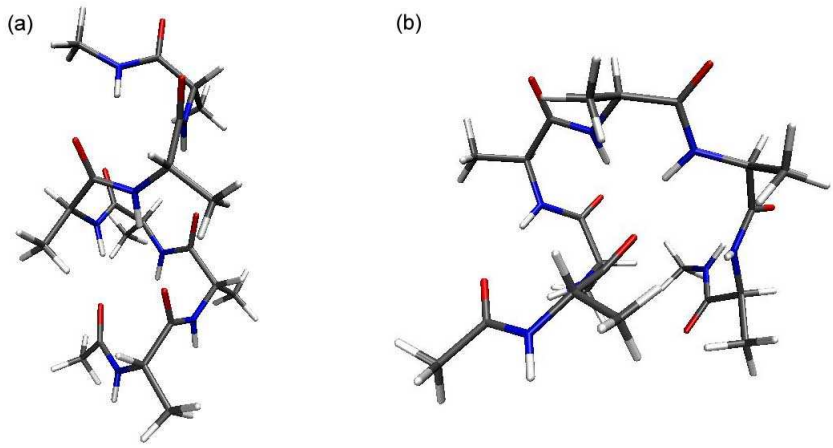


Figure 4. (a) Right-handed helical conformation of the alanine hexamer (global minimum). (b) Misfolded, extended, conformation of the alanine hexamer. All waters have been deleted for clarity.

References

1. A. Warshel and M. Levitt, *Theoretical studies of enzymic reactions – Dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme*, J. Mol. Biol., **103**, 227, 1976.
2. K. P. Eurenius, D. C. Chatfield, B. R. Brooks, and M. Hodoscek, *Enzyme mechanisms with hybrid quantum and molecular mechanical potentials: 1. Theoretical considerations*, Int. J. Quantum Chem., **60**, 1189, 1996.
3. Mulholland AJ, Lyne PD, and Karplus M, *Ab initio QM/MM study of the citrate synthase mechanism. A low-barrier hydrogen bond is not involved*, J. Am. Chem. Soc., **122**, 534, 2000.
4. A. Laio, J. VandeVondele, and U. Rothlisberger, *A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations*, J. Chem. Phys., **116**, 6941, 2002.
5. Y. Zhang, T. Lee, and W. Yang, *A pseudobond approach to combining quantum mechanical and molecular mechanical methods*, J. Chem. Phys., **110**, 46, 1999.
6. G. Alagona, P. Desmeules, C. Ghio, and P. A. Kollman, *Quantum mechanical and molecular mechanical studies on a model for the dihydroxyacetone phosphate blyceraldehyde phosphate isomerization catalyzed by triosephosphate isomerase (TIM)*, J. Am. Chem. Soc., **106**, 3623, 1984.
7. R. Car and M. Parrinello, *Unified approach for molecular dynamics and density functional theory*, Phys. Rev. Lett., **55**, 2471, 1985.
8. G. Galli and M. Parrinello, *The Ab-initio MD method*, Comp. Sim. Mat. Sci., **3**, 283, 1991.
9. D. Marx and J. Hutter, *Ab initio molecular dynamics: Theory and implementation*, Modern Methods and Algorithms for Quantum Chemistry, **1**, 301, 2000.

10. M. E. Tuckerman, *Ab initio molecular dynamics: Basic concepts, current trends and novel applications*, J. Phys. Condens. Matter, **14**, R1297, 2002.
11. G. J. Martyna and M. E. Tuckerman, *A reciprocal space based method for treating long range interactions in ab initio and force-field-based calculations in clusters*, J. Chem. Phys., **110**, 2810, 1999.
12. M. E. Tuckerman, P. Minary, K. Pihakari, and G. J. Martyna, *A new reciprocal space based treatment of long range forces on surfaces*, J. Chem. Phys., **116**, 5351, 2002.
13. P. Minary, J. A. Morrone, D. A. Yarne, M. E. Tuckerman, and G. J. Martyna, *Long range interactions on wires: A reciprocal space based formalism*, J. Chem. Phys., **121**, 11949, 2004.
14. D. Yarne, G. J. Martyna, and M. E. Tuckerman, *A Dual Space, Plane wave based approach to QM/MM calculations*, J. Chem. Phys., submitted (2001).
15. U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen, *A smooth particle-mesh Ewald method*, J. Chem. Phys., **103**, 8577, 1995.
16. T.A. Darden, A. Toukmaji, and L.G. Pedersen, *Long-range electrostatic effects in biomolecular simulations*, J. Chim. Phys., **94**, 1346, 1997.
17. E.L. Pollock and J. Glosli, *Comments on P(3)M, FMM, and the Ewald method for large periodic coulombic systems*, Comp. Phys. Comm., **95**, 93, 1996.
18. M. Patra, M. Karttunen, M. T. Hyvonen, E. Falck, P. Lindqvist, and I. Vattulainen, *Molecular dynamics simulations of lipid bilayers: Major artifacts due to truncating electrostatic interactions*, Biophys. J., **84**, 3636, 2003.
19. T. Laino, F. Mohamed, A. Laio, and M. Parrinello, *An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations*, J. Chem. Theory. Comput., **2**, 1370, 2006.
20. A. D. Becke, *Density-functional exchange-energy approximation with correct asymptotic behavior*, Phys. Rev. A, **38**, 3098, 1988.
21. C. Lee, W. Yang, and R. G. Parr, *Development of the Colle-Salvetti correlation-energy formular into a functional of the electron density*, Phys. Rev. B, **37**, 785, 1988.
22. D. K. Remler and P. A. Madden, *Molecular dynamics without effective potentials via the Car-Parrinello approach*, Mol. Phys., **70**, 921–951, 1990.
23. A. D. MacKerell, D. Bashford, M. Bellott, Jr. R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, III W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B, **102**, 3586, 1998.
24. N. Troullier and J. L. Martins, *Efficient pseudopotentials for plane-wave calculations*, Phys. Rev. B, **43**, 1993, 1991.
25. J. Gao, P. Amara, and M. J. Field, *A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations*, J. Phys. Chem. A, **109**, 4714, 1998.
26. B. M. Jonsson, H. Hakansson, and A. Liljas, *The structure of human carbonic anhydrase-II in complex with bromide and azide*, FEBS Lett., **322**, 186, 1993.
27. G. M. Torrie and J. P. Valleau, Chem. Phys. Lett., **28**, 578, 1974.
28. G. M. Torrie and J. P. Valleau, J. Comput. Chem., **23**, 187, 1977.
29. B. J. Berne, (Ed.), *Modern Theoretical Chemistry V*, Plenum, New York, 1977.
30. J. G. Kirkwood, J. Chem. Phys., **3**, 300, 1935.

31. E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, Chem. Phys. Lett., **156**, 472, 1989.
32. Michiel Sprik and Giovanni Ciccotti, *Free energy from constrained molecular dynamics*, J. Chem. Phys., **109**, 7737, 1998.
33. R. H. Swendsen and J. S. Wang, Phys. Rev. Lett., **57**, 2607, 1986.
34. K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn., **65**, 1604, 1996.
35. M. Tesi, E. J. J. Rensburg, E. Orlandini, and S. G. Whittington, J. Stat. Phys., **82**, 155, 1996.
36. P. Liu, B. Kim, R. A. Friesner, and B. J. Berne, Proc. Natl. Acad. Sci., **102**, 13749, 2005.
37. C. J. Woods, J. W. Essex, and M. A. King, *The development of replica-exchange-based free-energy methods*, J. Phys. Chem. B, **107**, 13703, 2003.
38. C. J. Woods, J. W. Essex, and M. A. King, *Enhanced Configurational Sampling in Binding Free-Energy Calculations*, J. Phys. Chem. B, **107**, 13711, 2003.
39. A. F. Voter, *Hyperdynamics: Accelerated molecular dynamics of infrequent events*, Phys. Rev. Lett., **78**, 3908, 1997.
40. A. F. Voter, *Parallel replica method for dynamics of infrequent events*, Phys. Rev. B, **57**, R13985, 1998.
41. F. G. Wang and D. P. Landau, *Efficient, multiple-range random walk algorithm to calculate the density of states*, Phys. Rev. Lett., **86**, 2050, 2001.
42. F. G. Wang and D. P. Landau, *Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram*, Phys. Rev. E, **64**, 056101, 2001.
43. J. I. Siepmann and D. Frenkel, *Configurational bias monte-carlo - a new sampling scheme for flexible chains*, Mol. Phys., **75**, 59, 1992.
44. Z. Zhu, M. E. Tuckerman, S. O. Samuelson, and G. J. Martyna, *Enhancing sampling of biomolecules using novel variable transformations*, Phys. Rev. Lett., **88**, 100201, 2002.
45. P. Minary, M. E. Tuckerman, and G. J. Martyna, *Dynamical spatial warping: A novel method for the conformational sampling of biophysical structure*, SIAM J. Sci. Comput., **30**, 2055, 2008.
46. A. Laio and M. Parrinello, *Escaping the free-energy minima*, Proc. Natl. Acad. Sci., **99**, 12562, 2002.
47. L. Rosso and M. E. Tuckerman, *An adiabatic molecular dynamics method for the calculation of free energy profiles*, Mol. Simul., **28**, 91, 2002.
48. L. Rosso, P. Minary, Z. Zhu, and M. E. Tuckerman, *On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles*, J. Chem. Phys., **116**, 4389, 2002.
49. L. Rosso, Jerry B. Abrams, and M. E. Tuckerman, *Mapping the backbone dihedral free-energy surfaces in small peptides in solution using adiabatic free-energy dynamics*, J. Phys. Chem. B, **109**, 4162–4167, 2005.
50. L. Maragliano and E. Vanden-Eijnden, *A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations*, Chem. Phys. Lett., **426**, 168, 2006.
51. Jerry B. Abrams, L. Rosso, and M. E. Tuckerman, *Efficient and precise solvation free energies via alchemical adiabatic molecular dynamics*, J. Chem. Phys., **125**, 074115,

2006.
52. C. Chipot and A. Pohorille, editors, *Free Energy Calculations*, Springer, Heidelberg, 2007.
53. ", The technique was first presented at a CECAM meeting on free energy calculations held in 2000 from June 19-21. The proceedings of this meeting were published in *Molecular Simulation*, Volume 28, Issues 1-2 (2002), which contains Ref.[47].
54. R. W. Zwanzig, J. Chem. Phys., **22**, 1420, 1954.
55. J. B. Abrams and M. E. Tuckerman, *Efficient and direct generation of multidimensional free-energy surfaces via aidabatic dynamics without coordinate transformations*, J. Phys. Chem. B, **112**, 15742, 2008.
56. E. Butkov, *Mathematical Physics*, Addison-Wesley, Reading, 1968.
57. Yi Liu and Mark E. Tuckerman, *Generalized Gaussian moment thermostatting: A new continuous dynamical approach to the canonical ensemble*, J. Chem. Phys., **112**, 1685–1700, 2000.
58. M. Tuckerman, B. J. Berne, and G. J. Martyna, *Reversible multiple time scale molecular-dynamics*, J. Chem. Phys., **97**, 1990, 1992.
59. G. J. Martyna, Mark E. Tuckerman, D. J. Tobias, and M. L. Klein, Mol. Phys., **87**, 1117, 1996.
60. R. J. Kennedy, K. Y. Tsang, and D. S. Kemp, ", 2002.
61. N. Kamiya, Y. S. Watanabe, S. Ono, and J. Higo, ", 2005.
62. G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello, *Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics*, J. Am. Chem. Soc., **128**, 13435, 2006.

QM/MM Methodology: Fundamentals, Scope, and Limitations

Walter Thiel

Max-Planck-Institut für Kohlenforschung
45470 Mülheim, Germany
E-mail: thiel@kofo.mpg.de

Combined quantum-mechanical/molecular-mechanical (QM/MM) methods have become a popular approach for modeling local electronic events in large systems with thousands of atoms. QM methods are used to describe the active site where chemical reactions or electronic excitations occur, and MM methods are employed to capture the effect of the environment on the active site. This article gives an overview over methodological and practical issues in QM/MM studies and outlines the scope and the limitations of current QM/MM applications.

1 Introduction

The QM/MM concept was introduced in 1976 by Warshel and Levitt who presented the first semiempirical QM/MM model and applied it to an enzymatic reaction[1]. The QM/MM approach found wide acceptance only much later, in the 1990s. Over the past decade, numerous reviews have documented the development of the QM/MM methodology and its application. Here we mention only a few of these[2–8] and refer to our own recent reviews[6,8] for an up-to-date coverage of the field with an extensive literature survey (755 and 627 references, respectively). The reader should consult these reviews for access to the original QM/MM papers since we shall quote only a small selection of these in the following.

The QM/MM approach is by now established as a valuable tool for modeling large biomolecular systems, but it is also often applied to study processes in explicit solvent and to investigate large inorganic/organometallic and solid-state systems. Methodological issues that are common to all these areas will be addressed in Sec. 2, while practical issues and potential pitfalls will be discussed in Sec. 3. Thereafter, an overview over QM/MM applications will be provided in Sec. 4. We conclude with a brief summary in Sec. 5.

2 Methodological Issues

The design of composite theoretical methods gives rise to a number of methodological problems that need to be solved. The basic idea is to retain (as much as possible) the formalism of the methods that are being combined and to introduce well-defined conventions for their coupling. In this section, we address the methodological choices that need to be made in the QM/MM case.

2.1 QM/MM partitioning

The entire system is divided into the inner QM region that is treated quantum-mechanically and the outer MM region that is described by a force field. There is a boundary region at the

interface where the standard QM and MM procedures may be modified or augmented in some way (e.g., by the introduction of link atoms or boundary atoms with special features, see below). The choice of the QM region is usually made by chemical intuition: one can normally define a minimum-size QM region on chemical grounds by considering the chemical problem at hand, and one can then check the sensitivity of the QM/MM results with respect to enlarging the QM region.

Standard QM/MM applications employ a fixed QM/MM partitioning where the boundary between the QM and MM regions is defined once and for all at the outset. It is also possible, but more involved, to allow the boundary to move during the course of a simulation (adaptive partitioning, "hot spot" methods) in order to describe processes with shifting active sites (e.g., the motion of solvated ions)[9].

QM/MM methods can be generalized from two-layer to multi-layer approaches, with a correspondingly extended partitioning. One such example is the use of a continuum solvation model as a third layer to mimic the effects of bulk solvent[10,11]. Other multi-layer approaches such as ONIOM go beyond the original QM/MM concept by integrating two or more QM regions[12].

2.2 Choice of QM method

The selection of a suitable QM method in QM/MM calculations follows the same criteria as in pure QM studies (accuracy and reliability versus computational effort). Traditionally, semiempirical QM methods have been most popular, and they remain important for QM/MM molecular dynamics (MD) where the computational costs are very high. Density functional theory (DFT) is the workhorse in many contemporary QM/MM studies, and correlated ab initio methods are increasingly used in electronically demanding cases or in the quest for high accuracy.

In small-molecule quantum chemistry, one nowadays often attempts to converge the results with regard to QM level and basis set. It has been demonstrated recently that this is also possible in QM/MM work on enzymes: using linear scaling local correlation methods the computed barriers for the rate-determining reactions in chorismate mutase and p-hydroxybenzoate hydroxylase (PHBH) can be converged to within 1–2 kcal/mol at the ab initio coupled cluster LCCSDT(0) level[13,14].

2.3 Choice of MM method

Established MM force fields are available for biomolecular applications (e.g., CHARMM, AMBER, GROMOS, and OPLS) and for explicit solvent studies (e.g., TIP3P or SPC for water). MM methods are generally less developed in other areas such as organometallic or solid-state chemistry which may pose restrictions on corresponding QM/MM work. Even in the favorable biomolecular case, it is often necessary to derive some additional force field parameters (whenever the QM/MM calculations target situations in the active-site region that are not covered by the standard force field parameters).

The classical biomolecular force fields contain bonded terms as well as nonbonded electrostatic and van der Waals interactions. Electrostatics is normally treated using fixed point charges at the MM atoms. The charge distribution in the MM region is thus unpolarizable which may limit the accuracy of the QM/MM results. The logical next step towards

enhanced accuracy should thus be the use of polarizable force fields which are currently developed by several groups in the biomolecular simulation community using various classical models (e.g., induced dipoles, fluctuating charges, or charge-on-spring models). The QM/MM formalism has been adapted to handle polarizable force fields[8,15], but one may expect corresponding large-scale QM/MM applications only after these new force fields are firmly established. In the meantime, essential polarization effects in the active-site environment may be taken into account in QM/MM studies by a suitable extension of the QM region (at increased computational cost, of course).

2.4 Subtractive versus additive QM/MM schemes

Subtractive QM/MM schemes are interpolation procedures. They require (i) an MM calculation of the entire system, (ii) a QM calculation of the inner QM region, and (iii) an MM calculation of the inner QM region. The QM/MM energy is then obtained simply by summing (i) and (ii) and subtracting (iii) to avoid double counting. In such an interpolation scheme, the QM/MM interactions are handled entirely at the MM level. This may be problematic with regard to the electrostatic interactions which will then typically involve fixed atomic charges in the QM and MM regions. Therefore, realistic MM parameters are also needed for the QM region which are often not available and difficult to obtain for typical QM/MM applications (where the QM region is "non-standard" and electronically demanding). These drawbacks have made subtractive QM/MM schemes less attractive, especially in the biomolecular area. On the positive side, it should be noted, however, that subtractive schemes are easy to implement and to generalize to the multi-layer case[12].

Additive schemes require (i) an MM calculation of the outer MM region, (ii) a QM calculation of the inner QM region, and (iii) an explicit treatment of the QM/MM coupling terms. The QM/MM energy is the sum of these three contributions. The coupling terms normally include bonded terms across the QM/MM boundary, nonbonded van der Waals-terms, and electrostatic interaction terms. The former two are generally handled at the MM level (using protocols that avoid double counting and related complications), while the latter one is modeled explicitly. This has the advantage that the electrostatic QM/MM interactions can be described realistically using QM-based treatments (see below). It is probably for this reason that the majority of the currently used QM/MM schemes are of the additive type.

2.5 Electrostatic QM/MM interactions

A hierarchy of models is available for handling the electrostatic coupling between the QM charge density and the MM charge model which may be classified[16] as mechanical embedding (model A), electrostatic embedding (model B), and polarized embedding (models C and D). They differ by the extent of mutual polarization between the QM and MM region.

Mechanical embedding is equivalent to the subtractive QM/MM scheme outlined above in that it treats the electrostatic QM/MM interactions at the MM level (typically between rigid atomic point charges). Both the QM and MM region are unpolarized in this case, and the QM charge density comes from a gas-phase calculation (without MM environment). This will often not be accurate enough, especially in the case of very polar environments (as in most biomolecules).

Electrostatic embedding allows for the polarization of the QM region since the QM calculation is performed in the presence of the MM charge model, typically by including the MM point charges as one-electron terms in the QM Hamiltonian. The electronic structure of the inner region can thus adapt to the environment, and the resulting QM density should be much closer to reality than that from a gas-phase model calculation. The majority of the current QM/MM work employs electrostatic embedding.

Polarized embedding attempts to capture the back-polarization of the MM region by the QM region as well, either in a one-way sense (model C) or in a fully self-consistent manner with mutual polarization (model D). The latter is the most refined embedding scheme which, however, has been applied only rarely up to now. It is expected to become more popular when general-purpose polarizable force fields are being used more often as MM components in QM/MM work, because polarized embedding is the natural coupling scheme in this case. As already mentioned above, polarization effects near the active site can alternatively also be taken into account with standard electrostatic embedding if the QM region is extended accordingly.

2.6 Boundary treatment

In many QM/MM studies it is unavoidable that the QM/MM boundary cuts through a covalent bond. The resulting dangling bond must be capped to satisfy the valency of the QM atom at the frontier, and in the case of electrostatic or polarized embedding, one must prevent overpolarization of the QM density by the MM charges close to the cut. To cope with these problems, there are essentially three different classes of boundary schemes that involve link atoms, special boundary atoms, and localized orbitals, respectively.

Link-atom schemes introduce an additional atomic center (usually a hydrogen atom) that is not part of the real system and is covalently bonded to the QM frontier atom. Each link atom generates three artificial nuclear degrees of freedom that are handled differently by different authors. The most common procedure is to fix the position of the link atom such that it lies in the bond being cut, at some well-defined distance from the QM frontier atom, and to redistribute the forces acting on it to the two atoms of the bond being cut (by applying the chain rule)[17]. This effectively removes the artificial degrees of freedom since the link-atom coordinates are fully determined by the positioning rule rather than being propagated according to the forces acting on them. Concerning the possible overpolarization in link-atom schemes, several protocols have been proposed to mitigate this effect which involve, for example, deleting or redistributing or smearing certain MM charges in the link region. Widely used is the charge-shift protocol[18].

Boundary-atom schemes replace the MM frontier atom by a special boundary atom that participates as an ordinary MM atom in the MM calculation, but also carries QM features to saturate the valency of the QM frontier atom in the QM calculation. These QM features are parametrized such that the boundary atom mimics the cut bond and possibly also the electronic character of the attached MM moiety. Examples for such schemes include the adjusted connection atoms for semiempirical QM methods[19], the pseudobond approach for ab initio and DFT methods[20], and the use of tailored pseudopotentials within plane-wave QM methods[21]. Properly parametrized boundary-atom schemes should be more accurate than link-atom schemes, but they are less popular in practice because the required special parameters are not generally available (only for selected bonds).

Localized-orbital schemes place hybrid orbitals at the boundary and keep some of them frozen such that they do not participate in the SCF iterations. These approaches are theoretically satisfying because they provide a boundary treatment essentially at the QM level. However, they are technically involved (mainly because of the orthogonality constraints that need to be imposed), and require transferability of the localized orbitals between model and real systems. Examples for such schemes are the local SCF method[22] in different variants[8] and the generalized hybrid orbital (GHO) method[23].

There have been several evaluations of and comparisons between the available boundary treatments. Overall the performance of link-atom schemes seems generally on par with localized-orbital approaches: both provide reasonable accuracy when applied with care. In practice, the link-atom scheme is most popular because of its simplicity and robustness, but the GHO treatment is also frequently used.

2.7 QM/MM geometry optimization

In theoretical studies of small molecules, potential energy surfaces (PES) are commonly explored by geometry optimization to locate the relevant stationary points (minima, transition states). This is also possible in QM/MM studies of large molecules with thousands of atoms, in principle, but it is obvious that one needs techniques that can handle thousands of degrees of freedom and are still efficient. The algorithms for manipulating coordinates should ideally be scaling linearly with the number of degrees of freedom, and the optimization should take advantage of the partitioning of the system into a QM region, where energy and gradient evaluation are computationally expensive, and an MM region, where these calculations are almost for free. Among the various approaches that have been proposed in this context[8], we only mention a linear-scaling fragment-based divide-and-conquer optimizer[24] and microiterative optimization strategies[25] with alternating geometry relaxation in the core region (containing the QM region) and the environment. Their combined use allows the efficient optimization of minima and transition states in large molecules at the QM/MM level even when using electrostatic or polarized embedding[26].

Given the vast configuration space that is accessible to the large molecules studied by QM/MM techniques, there are many closely related minima and transition states for any particular chemical reaction. QM/MM geometry optimizations of the stationary points along a single reaction path are therefore of limited significance. It is thus advisable in QM/MM optimization studies to determine at least several representative transition states with their corresponding minima in order to assess the influence of the conformational diversity of the environment; snapshots from classical MD simulations can serve as starting structures. Application of this procedure to the rate-limiting reaction in PHBH has shown that rms fluctuations of the computed QM/MM barriers for 10 snapshots are of the order of 2 kcal/mol[14]. Uncertainties of this magnitude must be anticipated when investigating only a single reaction path.

2.8 QM/MM molecular dynamics

The preceding discussion emphasizes the need for sampling configuration space in large molecules using molecular dynamics or related approaches. QM/MM MD calculations are computationally quite demanding, however, and routinely affordable only at the semiempirical QM/MM level. As in the case of QM/MM geometry optimization, this calls for

special techniques that reduce the computational cost by exploiting the QM/MM partitioning. One strategy is to avoid the expensive direct sampling of the QM region while fully sampling the MM configurations. An early example of this approach[27] kept the QM region fixed while sampling the MM region and used ESP(electrostatic potential)-derived charges for the QM atoms to evaluate the electrostatic QM/MM interactions during the MD run; this was shown to be successful in the context of a QM/MM free energy perturbation treatment in which the entropic contributions from the QM region are estimated separately[27,28]. There are a number of recent other activities to improve the available QM/MM MD technology[7,8].

2.9 QM/MM energy versus free energy calculations

Free energy differences govern chemical thermodynamics and kinetics, and theoretical studies should thus aim at free energy calculations. Statistical mechanics provides various techniques to determine free energy differences through sampling, e.g., thermodynamic integration, umbrella sampling, or free energy perturbation. All these techniques have been used in conjunction with semiempirical QM/MM methods in a straightforward manner[28–30], but they tend to become too expensive with ab initio or DFT QM components. For the latter case, approximate free energy treatments have been devised that have been reviewed recently[7].

In view of the computational effort and the technical difficulties of QM/MM free energy calculations, it is of interest to check how much the QM/MM results for energies and free energies differ in typical cases. There are not yet enough theoretical data available for a systematic assessment. However, judging from the QM/MM energy and free energy barriers for several enzymatic reactions, the differences often appear to be less than 1 kcal/mol for localized chemical events (e.g., hydrogen abstraction in cytochrome P450cam, OH transfer in PHBH, nucleophilic substitution in fluorinase, proton transfer in cystein protease). This confirms that the less demanding QM/MM geometry optimization studies can provide valuable information for many types of reactions.

3 Practical Issues

QM/MM calculations are not yet ”black-box” procedures. Therefore it seems worthwhile to address some of the practical problems and choices that are encountered in QM/MM work.

3.1 QM/MM software

QM/MM applications require efficient programs with wide-ranging functionality. Many of the commonly available QM and MM packages nowadays offer QM/MM capabilities as an add-on. The alternative is a modular approach that links external QM and MM codes via interfaces to a central core which supplies the QM/MM coupling as well as routines for standard tasks such as structure optimization, molecular dynamics, etc. The core also provides a common user interface to the external programs, at least for the most common options. The ChemShell software[31] is an example for such a modular QM/MM implementation which currently supports interfaces to several QM codes (GAUSSIAN,

TURBOMOLE, MOLPRO, ORCA, GAMESS-UK, NWChem, MNDO) and several MM force fields (CHARMM, GROMOS, AMBER, GULP).

When embarking on a QM/MM project it may be easiest to use the QM/MM capability of a standard QM or MM package that one is familiar with. In the long run, modular QM/MM software will offer more flexibility and allow the user to access more combinations of QM and MM methods and, in general, more QM/MM functionality.

3.2 QM/MM setup for biomolecular simulations

QM/MM studies on large systems such as enzymes require realistic starting structures. These will normally be derived from experiment (e.g., X-ray or NMR) because they cannot be generated by purely theoretical means. Small modifications of experimental structures are common in the setup phase, e.g., involving the replacement of an inhibitor by a substrate or the substitution of specific residues to generate the starting structure for a mutant of interest.

The available structural information from experiment is generally not complete and often not error-free. It thus needs to be checked and processed using the protocols that have been developed over the past decades by the classical simulation community. This involves, e.g., adding hydrogen atoms that are missing in X-ray structures, adding water molecules inside the biomolecule in "empty" spots, assigning the protonation states of titrable residues, and checking the orientation of residues in ambiguous cases. The system is then put into a water box and relaxed by a series of constrained energy minimizations and MD runs at the classical force field level; this may necessitate the derivation of force field parameters for the "non-standard" parts of the system. After equilibration, the system is subjected to a classical MD production run from which snapshots are taken as starting geometries for the QM/MM work. These starting structures typically contain the biomolecular system in a droplet of water (normally around 20000–30000 atoms).

It should be emphasized that this setup requires a lot of work prior to the actual QM/MM calculations. Errors and wrong choices (e.g., with regard to protonation states or the number of water molecules near the active site) cannot normally be recovered at a later stage. These issues have been discussed more thoroughly in a previous review[6], and further practical guidance is available in the original papers that deal with these questions[32,33]. Finally, while the preceding considerations have addressed the QM/MM setup for biomolecules, they should apply in an analogous manner to other systems with similar complexity.

3.3 Accuracy of QM/MM results

QM/MM calculations involve a lot of choices (see Sec. 2), and it is therefore very difficult to converge the QM/MM results with regard to all computational options. Typical biomolecular studies may employ DFT/MM calculations with a standard protein force field, electrostatic embedding, and a link-atom boundary treatment with a charge-shift scheme. The latter ingredients are considered as an integral part of the chosen QM/MM approach, and the sensitivity of the QM/MM results with regard to the chosen force field, embedding scheme, and boundary treatment is thus normally not checked (even though the QM/MM results will depend on these choices). On the QM side, different basis sets

are used in most DFT/MM studies to assess basis set convergence, and it is also common practice to check by how much the DFT/MM results change when using a different functional. Given the large computational effort in QM/MM work, it is not too surprising that high-level ab initio QM components are used rather seldom and that systematic convergence studies with respect to QM level and basis set are rare (unlike in small-molecule QM studies).

Conceptually, QM/MM treatments become more realistic upon extension of the QM region because the effects of the QM/MM coupling terms and of the MM force field on the active site should decrease by increasing the distance to the QM/MM boundary. It is thus highly advisable to validate the QM/MM results for any given application through QM/MM test calculations with larger QM regions.

3.4 QM/MM exploration of potential energy surfaces

In QM/MM geometry optimizations of systems with 20000–30000 atoms (see above) it is usually considered sufficient to allow only around 1000 atoms to move (i.e., the active site and the environment within a distance of typically 6–10 Å from the active site) while the outer part of the system remains fixed at the initially prepared snapshot geometry. This convention is beneficial in QM/MM studies of reaction profiles where it is essential to retain the same conformation of the optimized "active" region during the reaction in order to guarantee a smooth reaction path. Experience shows that this requirement can be well satisfied in practice with systems of around 1000 atoms, which becomes progressively more difficult for larger systems. If this requirement is not fulfilled (e.g., by the flip of a distant hydrogen bond or some other remote conformational change), the QM/MM results from geometry optimization become spurious since the PES is no longer smooth[32].

In QM/MM MD simulations of a large biomolecule in a water droplet, the outermost water layer is normally fixed or restrained such that there is no evaporation. Strict convergence criteria need to imposed in the QM part of the calculation to ensure energy conservation during the MD run[29]. Standard procedures can be applied to monitor the convergence of QM/MM MD simulations[29] and to analyze the results[30].

4 Applications

Biomolecular QM/MM studies constitute the largest application area, with enzymatic reactions as the prime target. Our previous reviews list 286 such QM/MM publications between 2001 and early 2006[6], and 179 such papers in the period 2006-2007[8]. A thorough survey of this work is obviously far beyond the scope of this article. Generally speaking, the QM/MM calculations provide detailed mechanistic insight into enzymatic reactions. The QM/MM energy, and particularly the QM/MM interaction energy, can be partitioned into its various components which offers the opportunity to analyze the effect of the protein environment (down to individual residues). Further insights can be gained by comparing the QM/MM results for the complete enzyme with QM results for suitably chosen model systems. In this manner, one can arrive at an improved understanding of the catalytic power of enzymes (as shown, for example, by a recent summary[8] of QM/MM studies on PHBH, chorismate mutase, and cytochrome P450).

QM/MM methods are suitable not only for studying chemical reactions in the active site of a large system, but also for investigating other localized electronic processes such as electronic excitation. In recent years there is an increasing number of QM/MM applications that address spectroscopic properties and electronically excited states. A typical procedure is to perform a DFT/MM geometry optimization or to extract snapshots from a semiempirical QM/MM MD run, followed by single-point calculations of spectroscopic properties at a suitable QM level (with inclusion of the MM point charges of the environment). QM/MM studies of this kind have been performed to compute not only electronic spectra (UV/vis absorption, emission, and fluorescence spectra), but also magnetic resonance spectra (NMR, EPR) and Mössbauer spectra. Examples include color tuning in the UV spectra of rhodopsins[34], NMR chemical shifts in rhodopsins[35] and in vanadium chloroperoxidase[36], as well as EPR and Mössbauer parameters in cytochrome P450cam[37]. QM/MM calculations can also be used to study excited-state reactivity in large systems (e.g., the photoisomerization in photoactive yellow protein[38] or the dynamics of a photoactive C–G base pair in DNA[39]).

Another QM/MM application area is experimental structure refinement of large biomolecular systems. The basic idea is to use a QM/MM, rather than a pure MM, model that is refined against the experimental data[40]. This is particularly advantageous in and around the active site since the standard biomolecular force fields are less reliable for the inhibitors or substrates that are present in this region. This approach has been applied to the refinement of X-ray, NMR, and EXAFS data[8].

The QM/MM applications outlined so far have been concerned with large biomolecules. As mentioned in the Introduction, QM/MM methods have also often been used to study processes in explicit solvent and in inorganic/organometallic and solid-state chemistry. An overview over these activities is beyond the scope of this article, leading references are available in our recent review[8].

5 Concluding Remarks

QM/MM methods are by now established as a powerful computational technique to treat reactive and other electronic processes in large systems. They can be applied whenever one needs to model a localized electronic event in an active site (typically of the order of 100 atoms) that is influenced by an interacting larger environment. Since they are not yet ”black-box” methods, one should exercise great care in the choice of the various computational QM/MM options and in the assessment of the results obtained. Despite the need to improve the available QM/MM tools further, especially with regard to higher accuracy and better sampling, there is a growing number of successful QM/MM applications in all branches of chemistry. This indicates that the existing QM/MM methods are good enough for the realistic modeling of real-world chemical problems.

Acknowledgments

This research was supported by the Max Planck Society. Many coworkers made essential contributions to our own QM/MM studies that have been mentioned in this article. Their names are listed in the references.

References

1. A. Warshel and M. Levitt, *Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme*, J. Mol. Biol. **103**, 227–249, 1976.
2. J. Gao, *Methods and applications of combined quantum mechanical and molecular mechanical potentials* in Reviews in Computational Chemistry, K. B. Lipkowitz and D. B. Boyd (Eds.), VCH, New York, Vol. **7**, 119-185, 1996.
3. P. A. Sherwood, *Hybrid quantum mechanics/molecular mechanics approaches* in Modern Methods and Algorithms of Quantum Chemistry, J. Grotendorst (Ed.), John von Neumann Institute for Computing, NIC Series Vol. **3**, 285-305, 2000.
4. J. L. Gao and D. G. Truhlar, *Quantum mechanical methods for enzyme kinetics*, Annu. Rev. Phys. Chem. **53**, 467–505, 2002.
5. R. A. Friesner and V. Guallar, *Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics methods for studying enzymatic catalysis*, Annu. Rev. Phys. Chem. **56**, 389–427, 2005.
6. H. M. Senn and W. Thiel, *QM/MM methods for biological systems* in Topics in Current Chemistry, M. Reiher (Ed.), Springer, Berlin, Vol. **268**, 173-290, 2007.
7. H. Hu and W. Yang, *Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods*, Annu. Rev. Phys. Chem. **59**, 573–601, 2008.
8. H. M. Senn and W. Thiel, *QM/MM methods for biomolecular systems*, Angew. Chem. Int. Ed. **48**, 1198–1229, 2009.
9. A. Heyden, H. Lin, and D. G. Truhlar, *Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations*, J. Phys. Chem. B **111**, 2231–2241, 2007.
10. P. Schaefer, D. Riccardi, and Q. Cui, *Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules*, J. Chem. Phys. **123**, 014905/1–14, 2005.
11. T. Benighaus and W. Thiel, *Efficiency and accuracy of the generalized solvent boundary potential for hybrid QM/MM simulations: Implementation for semiempirical Hamiltonians*, J. Chem. Theory Comput. **4**, 1600–1609, 2008.
12. M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, *ONIOM: A multilayered integrated MO + MM method for geometry optimizations and single point energy predictions*, J. Phys. Chem. **100**, 19357–19363, 1996.
13. F. Claeyssens, J. N. Harvey, F. R. Manby, R. A. Mata, A. J. Mulholland, K. E. Ranaghan, M. Schütz, S. Thiel, W. Thiel, and H.-J. Werner, *High accuracy computation of reaction barriers in enzymes*, Angew. Chem. Int. Ed. **45**, 6856–6859, 2006.
14. R. A. Mata, H.-J. Werner, S. Thiel, and W. Thiel, *Toward accurate barriers for enzymatic reactions: QM/MM case study on p-hydroxybenzoate hydroxylase*, J. Chem. Phys. **128**, 025104/1–8, 2008.
15. D. P. Geerke, S. Thiel, W. Thiel, and W. F. van Gunsteren, *Combined QM/MM molecular dynamics study on a condensed-phase S_N2 reaction at nitrogen: The effect of explicitly including solvent polarization*, J. Chem. Theory Comp. **3**, 1499–1509, 2007.
16. D. Bakowies and W. Thiel, *Hybrid models for combined quantum mechanical and*

molecular mechanical approaches, J. Phys. Chem. **100**, 10580–10594, 1996.
17. U. Eichler, C. M. Kölmel, and J. Sauer, *Combining ab initio techniques with analytical potential functions for structure predictions of large systems: Method and application to crystalline silica polymorphs*, J. Comp. Chem. **18**, 463–477, 1997.
18. P. Sherwood, A. H. de Vries, S. J. Collins, S. P. Greatbanks, N. A. Burton, M. A. Vincent, and I. H. Hillier, *Computer simulation of zeolite structure and reactivity using embedded cluster methods*, Faraday Discuss. **106**, 79–92, 1997.
19. I. Antes and W. Thiel, *Adjusted connection atoms for combined quantum mechanical and molecular mechanical methods*, J. Phys. Chem. A **103**, 9290–9295, 1999.
20. Y. Zhang, T.-S. Lee, and W. Yang, *A pseudobond approach to combining quantum mechanical and molecular mechanical methods*, J. Chem. Phys. **110**, 46–54, 1999.
21. A. Laio, J. van de Vondele, and U. Rothlisberger, *A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations*, J. Chem. Phys. **116**, 6941–6947, 2002.
22. V. Thery, D. Rinaldi, J. L. Rivail, B. Maigret, and G. G. Ferenczy, *Quantum mechanical computations on very large molecular systems: The local self-consistent field method*, J. Comp. Chem. **15**, 269–282, 1994.
23. J. Gao, P. Amara, C. Alhambra, and M. J. Field, *A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations*, J. Phys. Chem. A **102**, 4714–4721, 1998.
24. S. R. Billeter, A. J. Turner, and W. Thiel, *Linear scaling geometry optimisation and transition state search in hybrid delocalized internal coordinates*, Phys. Chem. Chem. Phys. **2**, 2177–2186, 2000.
25. F. Maseras and K. Morokuma, *IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states*, J. Comp. Chem. **16**, 1170–1179, 1995.
26. J. Kästner, S. Thiel, H. M. Senn, P. Sherwood, and W. Thiel, *Exploiting QM/MM capabilities in geometry optimization: A microiterative approach using electrostatic embedding*, J. Chem. Theory Comput. **3**, 1064–1072, 2007.
27. Y. Zhang, H. Liu, and W. Yang, *Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface*, J. Chem. Phys. **112**, 3483–3492, 2000.
28. J. Kästner, H. M. Senn, S. Thiel, N. Otte, and W. Thiel, *QM/MM free-energy perturbation compared to thermodynamic integration and umbrella sampling: Application to an enzymatic reaction*, J. Chem. Theory Comput. **2**, 452–461, 2006.
29. H. M. Senn, S. Thiel, and W. Thiel, *Enzymatic hydroxylation in p-hydroxybenzonate hydroxylase: A case study for QM/MM molecular dynamics*, J. Chem. Theory Comput. **1**, 494–505, 2005.
30. J. Kästner and W. Thiel, *Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: Umbrella integration*, J. Chem. Phys. **123**, 144105/1–5, 2005.
31. `http://www.chemshell.org`
32. A. Altun, S. Shaik, and W. Thiel, *Systematic QM/MM investigation of factors that affect the cytochrome P450cam-catalyzed hydrogen abstraction of camphor*, J. Comp. Chem. **27**, 1324–1337, 2006.
33. J. Zheng, A. Altun, and W. Thiel, *Common system setup for the entire catalytic cy-*

cle of cytochrome P450cam in quantum mechanical/molecular mechnical studies, J. Comp. Chem. **28**, 2147–2158, 2007.
34. M. Hoffmann, M. Wanko, P. Strodel, P. H. König, T. Frauenheim, K. Schulten, W. Thiel, E. Tajkhorshid, and M. Elstner, *Color tuning in rhodopsins: The mechanism for the spectral shift between bacteriorhodopsin and sensory rhodopsin II*, J. Am. Chem. Soc. **128**, 10818–10828, 2006.
35. J. A. Gascon, E. M. Sproviero, and V. S. Batista, *Computational studies of the primary phototransduction event in visual rhodopsin*, Acc. Chem. Res. **39**, 184–193, 2006.
36. M. P. Waller, M. Bühl, K. R. Geethalakshmi, D. Wang, and W. Thiel, *Vanadium NMR chemical shifts calculated from QM/MM models of vanadium chloroperoxidase*, Chem. Eur. J. **13**, 4723–4732, 2007.
37. J. C. Schöneboom, F. Neese, and W. Thiel, *Towards identification of the Compound I reactive intermediate in cytochrome P450 chemistry: A QM/MM study of its EPR and Mössbauer parameters*, J. Am. Chem. Soc. **127**, 5840–5853, 2005.
38. G. Groenhof, M. Bouxin-Cademartory, B. Hess, S. P. de Visser, H. J. C. Berendsen, M. Olivucci, A. E. Mark, and M. A. Robb, *Photoactivation of the photoactive yellow protein: Why photon absorption triggers a trans-to-cis isomerization of the chromophore in the protein*, J. Am. Chem. Soc. **126**, 4228–4233, 2004.
39. G. Groenhof, L. V. Schäefer, M. Boggio-Pasqua, M. Goette, H. Grubmüller, and M. A. Robb, *Ultrafast deactivation of an excited cytosine-guanine base pair in DNA*, J. Am. Chem. Soc. **129**, 6812–6819, 2007.
40. U. Ryde, L. Olsen, and K. Nilsson, *Quantum chemical geometry optimizations in proteins using crystallographic raw data*, J. Comp. Chem. **23**, 1058–1070, 2002.

DFT Embedding and Coarse Graining Techniques

James Kermode[1,2]**, Steven Winfield**[1]**, and Gábor Csányi**[2]**, and Mike Payne**[1]

[1] TCM Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge, CB3 0HE, UK
E-mail: jrk33@cam.ac.uk

[2] Engineering Laboratory, Trumpington Street, Cambridge, CB2 1PZ, UK

Classical molecular dynamics and first principles quantum mechanical calculations are two of the most important methods currently used to model physical systems at the atomic level. The former allow simulations of millions of atoms to be carried out on a nanosecond timescale but the accuracy is limited by the requirement to use simple parameterisations as interatomic potentials. If the scientific question of interest can be effectively answered by considering the behaviour of a very small number of atoms, up to around a hundred, then *ab initio* approaches allow this limitation to be overcome. In many cases we can extract enough information from these accurate quantum mechanical calculations to parameterise less transferable, but far less expensive, models and use them on a larger length scale. For some systems however, it is impossible to separate the behaviour on the various length scales, since the coupling between them is strong and bidirectional. Then the only option is to carry out a *hybrid* simulation, where some parts of the system are treated at a higher level of accuracy; this is the subject of this lecture.

1 Introduction

Over the last twenty years, the *ab initio* methods described in the previous lectures have made modelling of simple systems reliable, accurate and routine. This is partly due to the significant increase in capacity and speed of available computers and partly to the development of high quality codes that make effective use of these resources. As a result,

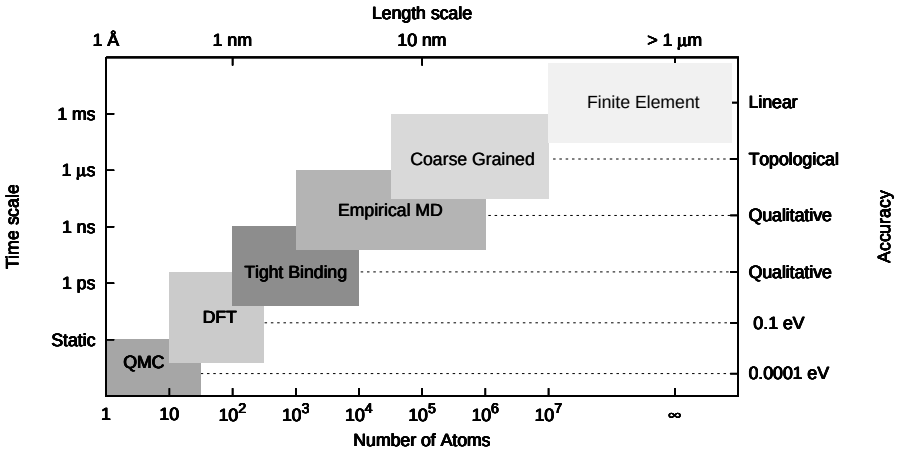


Figure 1. Schematic representation of the range of length- and time-scales accessible to a variety of modelling methods, from quantum Monte Carlo (QMC) for very accurate, very expensive static calculations through to approximate methods such as finite-element modelling.

attention is now focusing on modelling larger scale, more complex systems. Representative examples from fields as disparate as biology and materials science include enzyme catalysis[1] and defect migration in semiconductors. Ideally, we would simulate such systems using entirely first principles methods, free of empirical parameters and the accuracy and transferability problems associated with them. However, *ab initio* molecular dynamics is limited to simulating a few hundred atoms for up to a few picoseconds. Using more approximate methods (*e.g.* tight binding), the number of atoms can be extended to perhaps thousands, or the time period increased by a few orders of magnitude, but for many problems this is still insufficient. Fig. 1 illustrates the approximate range of application of various modelling techniques and makes clear the challenges we face if we wish to model complex problems with high accuracy.

1.1 Hierarchical Modelling

In recent years, there has been a great deal of work on multiscale methods that attempt to apply accurate quantum calculations to larger systems in one way or another. Most commonly, such methods are examples of *hierarchical multiscale modelling*, where the results of a calculation at one scale are used to parameterise less accurate calculations at a larger scale, making bigger systems or longer simulation times possible. There are many such examples of DFT based coarse graining in the literature including: using DFT forces to parameterise empirical interatomic potentials; calculating defect energies, often in different charge states, to their determine equilibrium concentrations; calculating surface stresses and the energies of surface steps to determine the thermodynamic properties of stepped surfaces.

For many materials and biological processes, the relevant timescale is of the order of milliseconds or longer, well beyond the capability of traditional molecular dynamics. To make progress we can either form a hierarchical multiscale model by coarse graining the system and considering the dynamics of the aggregate particles, or we can try to extract activation energies and reaction pathways from static calculations or short MD runs to parameterise Monte Carlo models. For a review of hierarchical multiscale methods and examples of their application, see Ref. 2.

1.2 Simultaneous Modelling

There is a large class of problems where the physical processes on the various length scales are strongly coupled and cannot be separated into a series of independent calculations; often this is because the nanoscale phenomena is driven by forces determined at least partially on the macroscopic scale. Simulation of such systems requires *simultaneous* coupling of length scales. Over the last ten years there has been much effort to devise schemes, referred to as *hybrid* or *embedded* methods, that combine a range of modelling techniques into a single simulation.

Occasionally, the large scale processes are so simple that they can be simulated very easily, as an example Martonak[3] added a classical pressure reservoir of soft spheres to an *ab initio* simulation of a small molecule. Usually, however, the large scale behaviour requires a more complex model to accurately capture the physics; this will be assumed to be the case for the remainder of the work discussed in the lecture.

1.3 Multiscale Applications in the Solid State

Stress induced defect processes in metals and semiconductors often give rise to strongly coupled multiscale phenomena. Examples include point-defect diffusion, dislocation motion, grain boundaries and, of course, the prototypical multiscale modelling problem: fracture.

Point-Defect Diffusion The stability and migration of point defects in semiconductors is affected both by local chemical interactions and long range strain fields. An example long range effect is the strain field resulting from the lattice mismatch between epitaxial layers in semiconductors. Although the quantum mechanical treatment of the bonding rearrangement around a defect requires only a few hundred atoms, we would need to include thousands more atoms to accurately represent the inhomogenous strain environment, particularly if we are to model interactions between multiple defects.

Dislocation Motion The strength of many materials is dominated by the behaviour of their dislocations. The core of a dislocation is a 1D region in which the bonding is significantly distorted. Dislocations in covalent materials move by the formation of kinks in the dislocation, where the bonding is very highly distorted. As the kink moves, so does this region of distortion. This motion requires bond breaking and reformation, therefore this region should be modelled by a highly accurate quantum mechanical technique.

Grain Boundaries It is not always possible to assume perfect single crystal structure and ignore the effect of grain boundaries when studying the physical and electronic properties of semiconductors. This is true for many materials of growing technological relevance, for example gallium nitride, silicon carbide and diamond.[4] Grain boundaries change the crystal structure on two length scales: they introduce long-range elastic distortion and local bonding disorder. They also act as sinks and sources for dislocations and traps for dopants, electrons and holes, further increasing the local chemical complexity. A multiscale description is needed to describe these systems since empirical interatomic potentials describe the long range-interactions adequately but local rebonding requires quantum mechanical accuracy.

Brittle Fracture Fracture is perhaps the best example of a multiscale materials process. The conditions for crack propagation are created by stress concentration at the crack tip, and depend on macroscopic parameters such as the loading geometry and dimensions of the specimen.[5–8] In real materials, however, the detailed crack propagation dynamics, are entirely determined by atomic scale phenomena since brittle crack tips are atomically sharp and propagate by breaking bonds, one at a time, at each point along the crack front.[9,10] This means the tip region is primarily a one dimensional line, perpendicular to the direction of propagation, and so it should be possible to define a contiguous embedding region to be treated with a more accurate model in a hybrid simulation. There is a constant interplay between the length scales because the opening crack gives rise to a stress field with a singularity at the tip,[11] as illustrated in Fig. 2, and in turn it is this singular stress field which breaks the bonds that advance the crack. Only by including the tens of thousands of atoms that contribute significantly to the elastic relaxation of this stress field can we hope to accurately model the fracture system, and thus a multiscale approach is essential.

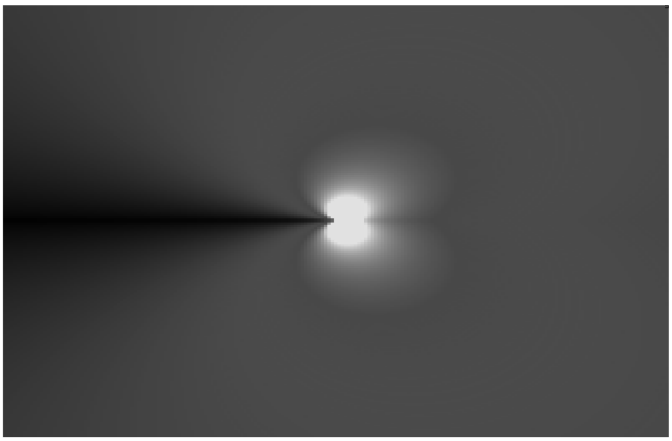


Figure 2. Maximum principal stress near the the tip of a crack under uniaxial tension in the opening mode, from the linear elastic solution. Light areas are the most highly stressed and dark the least.

Brittle fracture is the prototypical problem that has spurred many recent advances in the field of hybrid modelling of materials systems. For an example of a recent hybrid approach to modelling the fracture of silicon, see Ref. 12.

2 Coupling Continuum and Atomistic Systems

In the lecture, we shall concentrate on hybrid schemes which link quantum mechanical and classical modelling, but to provide some historical background, we shall first look briefly at a larger length scale. The pioneering hybrid simulations of materials systems were performed by Kohlhoff,[13] where classical atomistic and continuum elastic models were coupled to successfully model the directional cleavage anisotropy of a BCC crystal. This approach has been developed in the *quasicontinuum* (QC) method of Tadmor *et al.*[14]

The key problem with coupling atomistic and continuum models of matter is finding ways to connect these conceptually very different descriptions. Atomic positions need to be mapped onto a continuous displacement field, and energy calculations from interatomic potentials in the atomistic region and constitutive laws in the continuum region need to be harmonised. In the QC approach, a small subset of the atoms that would appear in a fully atomistic model are selected to represent the system as a whole, with a higher sampling density in highly deformed regions. The system is divided into cells, with one representative atom in each cell, as illustrated in Fig. 3. We assume that the energy of all the atoms in each cell is the same as that of the representative atom. The energies of these representative atoms are computed from the local environment, either from constitutive laws in areas that are nearly homogeneously deformed, or fully atomistically for non-uniformly deformed regions.

The atomistic and continuum methods are not completely compatible: non-physical forces arise on the continuum side of the boundary since it looks like an artificial surface.

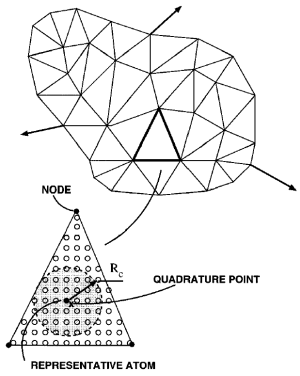


Figure 3. Schematic illustration of the finite element discretisation of a solid in the quasicontinuum method. The lower panel shows the representative atom for a particular triangular element. Reproduced from Ref. 14

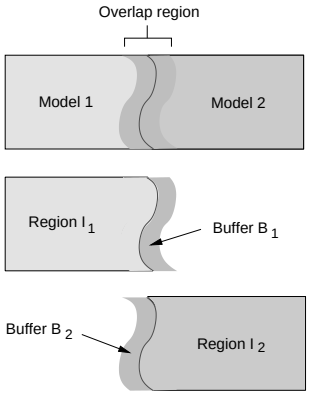


Figure 4. Schematic showing how overlap buffers can be used to solve the boundary problem in a classical/classical embedding scheme, where all interactions are short ranged.

The atomistic interactions used in QC are limited to be nearest neighbour models so there are no artificial forces in the atomistic region. In a refinement of the QC method, Shenoy *et al.*[15] removed these ghost forces in what they called the dead load approximation. The QC method has been applied to many systems, for example to study the interaction of dislocations with grain boundaries[16] and the effect of grain orientation on fracture.[17]

3 Coupling Two Classical Atomistic Systems

As a prelude to looking at the difficulties posed when attempting to couple quantum and classical systems, let's consider how two classical atomistic models could be combined. Providing the models are both short ranged, a straightforward treatment of the boundary is possible. We allow the regions to overlap as shown in Fig. 4, then evaluate the energy for the two regions separately, with a buffer region for both calculations. The locality of

the classical potentials means that each of these energies is a sum of local energies ϵ_i for each atom, so it is easy to separate the energy into a contribution due to the interior atoms and one due to the buffer atoms. For example, the energy for Model 1 of Fig. 4 can be decomposed as

$$E^{(1)} = \sum_i \epsilon_i = \underbrace{\sum_{i \in I_1} \epsilon_i}_{E_I^{(1)}} + \underbrace{\sum_{i \in B_1} \epsilon_i}_{E_B^{(1)}} \tag{1}$$

where I_1 and B_1 denote the interior and buffer sections of region one, as shown in Fig. 4. The same decomposition can be applied to give $E^{(2)}$ for Model 2. The total hybrid energy is then obtained by summing the contributions from the two interior regions, neglecting the buffers:

$$E_{\mathrm{hybrid}} = E_I^{(1)} + E_I^{(2)} \tag{2}$$

The artificial surfaces created at the boundary will be much more of a problem when we come to consider embedding a non-local quantum system which is described in the next section.

4 Coupling Quantum and Classical Systems

Coupling quantum and classical systems poses significantly greater challenges than combining two classical descriptions of matter. As quantum mechanics is non-local the simple partitioning scheme described above will not work. To overcome this problem we need to provide appropriate boundary conditions for the quantum calculations and find a way to spatially localise their effects.

The quantum mechanical model is assumed to be accurate enough to describe the physics of the region of interest correctly, perhaps using tight binding or an *ab initio* approach. The classical model needs only to correctly capture the basic topology of bonding and give the correct response to small elastic deformations, while remaining inexpensive to compute: empirical interatomic potentials are ideal for this purpose. Furthermore, since we shall use the quantum model anywhere we suspect the classical model will be unreliable, we prefer that the classical model be robust and inexpensive rather than being highly transferable. There has been a great deal of effort in recent years to produce potentials which attempt to model complex processes such as defect formation — generally we have found that such potentials are not useful in a hybrid simulation. We prefer simple potentials such as the Stillinger-Weber model to more complex ones such as EDIP.

The widely used assumption, upon which all quantum/classical hybrid schemes rely, is that the physics is local so that observables can be computed locally, taking into account only atoms which lie within some finite distance of the region of interest. Equivalently, we require that distant trajectories are instantaneously independent. Providing the quantum region is large enough, the trajectories that are important are not affected by the fact that, far away, the system is treated classically. However, it is also necessary, from a practical point of view, that the quantum trajectories can be computed accurately using a small quantum region. Both these conditions are satisfied by the *strong locality* condition:

$$\frac{\partial^n}{\partial \mathbf{r}_j^n} \frac{\partial E_{\mathrm{total}}}{\partial \mathbf{r}_i} \to 0 \ \text{ as } \ |\mathbf{r}_i - \mathbf{r}_j| \to \infty \ \ \forall\, n \in \mathbb{N}, i \neq j \tag{3}$$

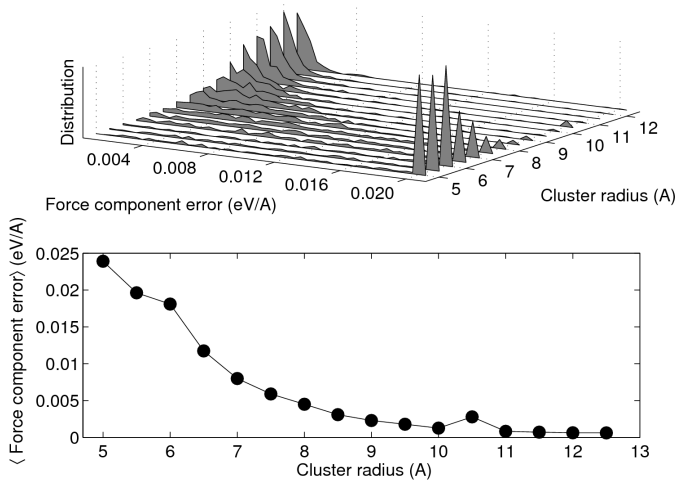


Figure 5. Top: distribution of force component errors at the centre of a finite cluster for a range of cluster sizes. The peaks on the right represent the integrated distribution of errors which are larger than 0.02 eV Å^{-1}. The system is a silicon self-interstitial in a 512-atom cubic cell with periodic boundary conditions, equilibrated at 1400 K. Bottom: the mean of the absolute error in the force components as a function of cluster radius. The peak at 10.5 Å is due to the periodicity of the underlying system. Clusters were terminated with hydrogen atoms. The force model is tight binding, from Ref.[18]. This figure is reproduced from Ref.[19].

where $\mathbf{r}_i$ and $\mathbf{r}_j$ are the positions of atoms i and j. This spatial localisation of observables is a stronger requirement than that the density matrix be sparse so that its elements decay rapidly as the separation between two atoms increases. The strong locality assumption can be tested for a particular system by testing the rate of convergence of the force on the central atom of a cluster as the cluster radius is increased. Fig. 5 shows an example of a test of strong locality for silicon using a tight binding force model. Most quantum systems either obey strong locality, or at least the parts of the Hamiltonian that do not, such as long range Coulomb and van der Waals interactions, can be dealt with in a purely classically manner.

Before we consider the details of the coupling strategy, it is appropriate to ask what we want from an ideal hybrid simulation. It is not feasible for the atoms in the quantum region to move as if the whole system were described quantum mechanically, since the classical atoms still move along classical trajectories, and the quantum atoms will respond to the new positions of the classical atoms. Hence, the best we can aim for is for the quantum atoms to behave *instantaneously* as if they are embedded in a fully quantum system.

5 The QM/MM Approach

The earliest quantum/classical hybrid simulation was performed by Warshel[1] in 1976 in which they model the reaction of the enzyme lysozyme. Enzyme catalysis is often controlled by large scale motion of macromolecules, with a small active site at which the chemical reaction takes place. Warshel and Levitt noted the need to describe the active site at a quantum mechanical level of detail to give an accurate description of hydrophobic

interactions and hydrogen bonding during catalysis. The electrostatic environment at the active site is determined by the configuration of the entire system of enzyme, substrate and solvent. The long range nature of electrostatic forces in such systems means that atoms far from the active site respond to the presence of a substrate, which in turn causes a change in the local electrostatic environment.

Hybrid methods of this kind, where the dominant interaction is electrostatic, have become known as quantum mechanical/molecular mechanical (abbreviated QM/MM) methods. They have become very popular in the biological and biochemical modelling communities in recent decades. All the fundamental aspects of modern QM/MM techniques were contained in Warshel's pioneering work: the quantum region was chosen very carefully by hand, and the boundary atoms were terminated with a frozen hybrid orbital.

In this section we will give only a brief overview of the QM/MM method which was covered in detail in Professor Thiel's lecture. In the context of the present lecture, DFT embedding should be understood simply as using DFT as the 'QM' method in a QM/MM scheme. It is worth pointing out that compared to quantum chemistry methods, such as Hartree Fock, DFT has a significant disadvantage for embedding schemes. It is known that DFT underestimates bandgaps. One consequence is that the density matrix is less localised in DFT than in quantum chemistry approaches and so DFT embedding is expected to be more sensitive to boundary effects. For reviews of recent developments in the QM/MM field see Ref. 20 and Ref. 21. Within the QM/MM framework, the total energy is the sum of three contributions: the quantum mechanical energy of the quantum region, the classical energy of the rest of the system, and a term representing the interaction between the two. There are two distinct approaches to performing a QM/MM calculation, which differ in their treatment of the interaction between regions: *mechanical embedding* and *electrostatic embedding*. We describe each of these below.

5.1 Mechanical Embedding

Mechanical embedding schemes perform quantum calculations for the QM region in the absence of the MM region, treating the interactions between the regions classically. The simplest mechanical embedding scheme is the two-layer ONIOM method,[22] illustrated in Fig. 6. Here the total energy is obtained from three independent calculations:

$$E_{QM/MM} = E_{QM}(QM) + E_{MM}(QM + MM) - E_{MM}(QM) \tag{4}$$

where the subscripts denote the energy model and the function arguments indicate the parts of the system to be included in each calculation. The MM system contains all the atoms and the quantum system contains the atoms of quantum mechanical interest plus *link atoms* used to cap dangling covalent bonds. ONIOM relies on cancellation of errors between the two surface energies.

There are two major drawbacks to the mechanical embedding approach. Firstly it is not always possible to obtain an accurate set of electrostatic MM parameters for atoms in the QM region; this is a particular problem since it is often the unavailability of such parameters which motivates the desire to treat this region quantum mechanically in the first place. Secondly, the scheme ignores perturbations in the electronic structure of the QM region caused by the charge distribution of the MM region. The three layer ONIOM method[23,24] goes some way to solving these problems by introducing an intermediate layer,

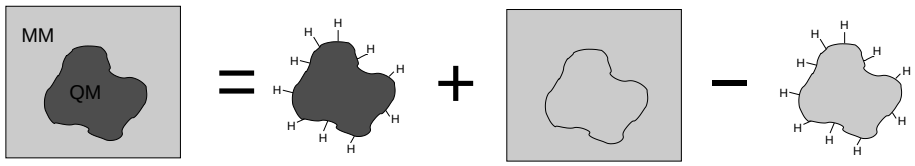


Figure 6. The two-layer ONIOM-style QM/MM scheme. Dark regions are treated with QM and light regions with MM. The termination atoms indicated as H could in fact be a more complex pseudoatom termination.

treated at semi-empirical quantum mechanical accuracy. This allows a consistent treatment of the polarisation of the active site.

5.2 Electrostatic Embedding

In an electrostatic embedding scheme the QM calculation is carried out in the presence of the classical charge distribution by adding terms that describe the electrostatic interaction between regions to the QM Hamiltonian. Normally atom centred partial point charges are used, but more advanced techniques employ a multipole expansion of the electric field for increased accuracy. Bonded and van der Waals interactions between the regions are still treated classically.

One problem of the standard electrostatic embedding approach is that classical atoms just outside the quantum region look appear as bare coulomb charges in the quantum calculation. There is a tendency for electron density to unphysically 'spill-out' onto these atoms to neutralise these charges. Laio *et al.*[25] have developed an efficient implementation of an electrostatic embedding scheme that addresses this issue by dealing with the short and long range electrostatic interactions differently to avoid spill-out.

5.3 Termination of Covalent Bonds

The QM/MM method has been applied fairly extensively to multiscale solid state systems of the types described at the beginning of this chapter. Electrostatic screening is very effective in metals and small band gap insulators, so mechanical embedding schemes are widely used for such systems. Bonded interactions between the QM and MM regions are much more of a problem in the solid state, since for a typical spherical QM region the number of covalent bonds that have to be cut to generate the QM cluster is of the same order as the number of atoms in the region.

To incorporate a quantum mechanical calculation of a subsystem into the total Hamiltonian, these artificially cut bonds must be terminated. There are various methods for doing this, usually based on using hydrogen link atoms or parameterised semiempirical 'pseudoatoms' that attempt to mimic the electronic effect of the region outside the subsystem that has been removed. A localised orbital parameterised with calculations on small model systems can be used to provide a quantum mechanical description of the charge distribution around the QM/MM boundary.[26] This approach is less widely used since is not possible to include these hybrid orbitals in a plane wave *ab initio* code.

We have seen in Section 2 that these termination strategies are sufficient to give accurate classical forces, since the classical description of covalent bonding is very near

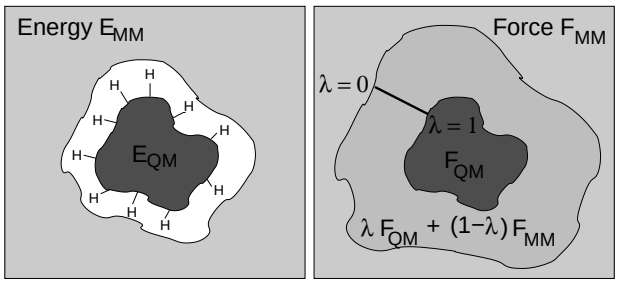


Figure 7. Comparison of QM/MM (left panel) and force mixing (right panel) approaches to the electronic termination problem.

sighted. Quantum mechanics, however, is not a nearest neighbour model, so atoms close to the terminated boundary of the QM subsystem feel an artificial environment, no matter how complex the passivation scheme employed. Moreover, it is impossible to exclude the contribution of the termination atoms to the total quantum mechanical energy of the subsystem. In a typical covalent system the length scale for strong locality of the electronic energy is the order of a nanometre, so any termination method that merely replaces a broken bond with a single atom cannot hope to give accurate forces at the boundary, due to the non-local nature of the quantum mechanical forces.

5.4 Force Mixing

An alternative to the standard QM/MM termination method is to move smoothly from quantum to classical forces over a transition region. This is the *force mixing* technique, where the forces used for the dynamics are interpolated, commonly linearly according to

$$\mathbf{F} = \lambda \mathbf{F}_{\mathrm{QM}} + (1 - \lambda)\mathbf{F}_{\mathrm{classical}} \tag{5}$$

with λ varying from zero on the classical edge of the transition zone to one at the QM edge. Higher order interpolation is also possible. A comparison of traditional QM/MM termination and force mixing is illustrated in Fig. 7.

Compared to the link atom method, force mixing slightly reduces the effect of inaccurate forces on atoms near to the edge of the QM region, since they are reduced in weight and mixed with classical forces. However, since the strong locality length scale is large, the transition zone must be very wide for this to have much of an effect, so large quantum mechanical zones are required.

A major disadvantage of force mixing is that since the forces no longer come from a single Hamiltonian neither energy nor momentum are conserved. The resulting dynamics can be unphysical. The action-reaction principle is not obeyed so, for example, the forces on a dimer spanning the boundary do not sum to zero. This creates a mechanical incompatibility across the boundary, which can lead to instabilities in the dynamics. Nevertheless, force mixing continues to be the most widely used approach for hybrid simulation of solid state systems.

The earliest quantum/classical multiscale fracture simulations were published in 1993 by Spence.[27] They describe their approach as a flexible boundary condition for an *ab initio*

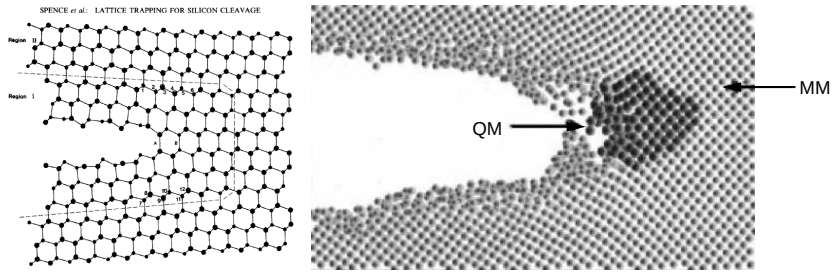


Figure 8. Early hybrid fracture simulation techniques. Left panel: relaxed 324 atom obtained using flexible boundary conditions. Region I (120 atoms) was treated with an *ab initio* method and Region II with an interatomic potential. Reproduced from Ref. 27. Right panel: snapshot from a MAAD simulation of fracture, showing the decomposition of the simulation into finite element (FE), molecular dynamics (MD) and tight binding (TB) regions. Reproduced from Ref. 28.

calculation, but it is effectively a force mixing embedding scheme. Alternate relaxations of the two regions illustrated in Fig. 8a were performed, with an overlap buffer to ensure self consistency. Some years later, Broughton[28] proposed the MAAD (macroatomistic *ab initio* dynamics) method which couples finite elements, molecular dynamics and semiempirical tight binding in a QM/MM approach to model crack propagation; a snapshot of the dynamics is shown in Fig. 8b. Pseudoatom terminator 'silogens' designed to behave like monovalent silicon atoms were used to terminate the tight binding region and a force mixing embedding approach was used.

Ogata's group has applied the ONIOM method to the simulation of cracks,[29] surface oxidation[30], and more recently they have investigated the effect of water on the initiation of corrosion induced cracks.[31] The group uses an improved version of ONIOM called the buffered-cluster method.[32] The QM region is cut out as normal, but then buffer atoms are added to terminate broken covalent bonds. The buffer is then relaxed using the classical force model, resulting in a relaxed buffered cluster which gives better surface error cancellation since it is closer to the equilibrium bulk structure.

5.5 Multiple Layer Termination

In 2001, Bernstein and Hess[33] proposed a modified treatment of the quantum zone boundary that addresses the electronic termination problem. They used a Green's function technique to create a *transition zone* with a thickness of several atomic layers which is included in the quantum mechanical calculation. Forces from this zone are not included in the dynamics. This method was later employed in a hybrid classical and tight binding simulation,[34] referred to as the DCET (dynamic coupling of empirical potential and tight binding) method. The combination of transition zone and force mixing gives accurate quantum mechanical forces and allowed the QM region to be moved during a simulation for the first time. However, the force mixing technique requires a large QM region, making the method difficult to scale up to a full *ab initio* calculation.

There is an alternative, more straightforward, termination strategy: we can obtain accurate *forces* for all atoms in the quantum region by using a wider buffer region. If we include a thick enough shell of nominally classical atoms in the quantum calculation, then

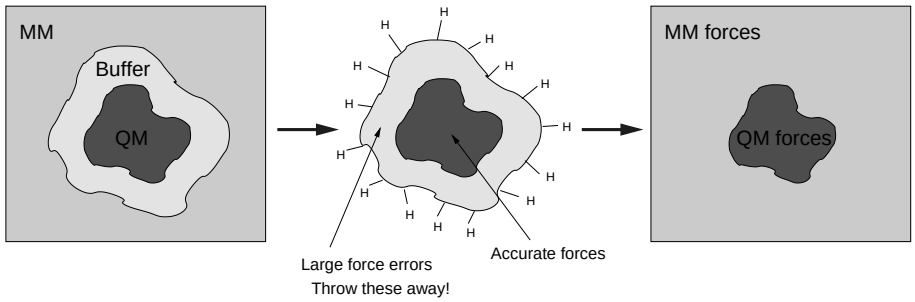


Figure 9. The finite buffer termination strategy. Forces on the atoms in the buffer region are discarded to give accurate hybrid forces on all atoms (right hand side).

the forces on the QM atoms themselves will be accurate. Since these forces are local quantities, we can easily discard the contaminated termination region and keep only the forces on the original QM atoms. This finite buffer scheme, illustrated in Fig. 9, is a major ingredient of the 'Learn on The Fly' hybrid method which will be the primary topic of the lecture.

These multiple layer termination approaches can solve the electronic termination problem, but there will still be a mechanical incompatibility across the boundary. If we used forces from the finite buffer scheme to do molecular dynamics, the resulting trajectories could be unstable, exactly as in the force mixing approach discussed above.

6 Summary

This lecture has introduced hybrid modelling techniques and reviewed a number of approaches which allow simultaneous simulation of coupled quantum and classical systems. We have seen that the fundamental difficulty of constructing such a hybrid modelling scheme lies in finding an effective treatment of the boundary. This is a particular problem in solid state systems, where many covalent bonds have to be cut to form the QM cluster and this has restricted the application of the QM/MM method. The problem can be divided into the electronic termination problem, which can be solved by discarding the inaccurate forces in a buffer zone at the edge of the QM region, and the mechanical matching problem. A solution to this mechanical mismatch is the basis of the 'Learn on The Fly' method.

References

1. A. Warshel and M. Levitt, *Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme*, J. Mol. Biol., **103**, 227–249, 1976.
2. Risto M Nieminen, *From atomistic simulation towards multiscale modelling of materials*, J. Phys.: Cond. Mat., **14**, no. 11, 2859–2876, 2002.

3. R. Martoňák, C. Molteni, and M. Parrinello, Ab Initio *Molecular Dynamics with a Classical Pressure Reservoir: Simulation of Pressure-Induced Amorphization in a $Si_{35}H_{36}$ Cluster*, Phys. Rev. Lett., **84**, no. 4, 682–685, Jan 2000.
4. Maki A. Angadi, Taku Watanabe, Arun Bodapati, Xingcheng Xiao, Orlando Auciello, John A. Carlisle, Jeffrey A. Eastman, Pawel Keblinski, Patrick K. Schelling, and Simon R. Phillpot, *Thermal transport and grain boundary conductance in ultrananocrystalline diamond thin films*, J. Appl. Phys., **99**, no. 11, 114301, 2006.
5. C. E. Inglis, *Stresses in a plate due to the prescence of cracks and sharp corners*, Trans. Inst. Naval. Archit., **55**, 219, 1913.
6. A. A. Griffith, *The Phenomena of Rupture and Flow in Solids*, Philos. Trans. R. Soc. London A, **221**, 163, 1921.
7. K. B. Broberg, *Cracks and Fracture*, Academic Press, San Diego, CA, 1999.
8. L. B. Freund, *Dynamic Fracture Mechanics*, Cambridge Univ. Press, Cambridge, UK, 1990.
9. B. Lawn, *Fracture of Brittle Solids — Second Edition*, Cambridge Univ. Press, Cambridge, UK, 1993.
10. I. H. Lin and R. Thomson, *Cleavage, Dislocation Emission, and Shielding for Cracks Under General Loading*, Acta Metall., **34**, 187–206, 1986.
11. G. R Irwin, *Fracturing of Metals*, pp. 147–166, American Society for Metals, Cleveland, OH, 1948.
12. J. R Kermode, T Albaret, D Sherman, N Bernstein, P Gumbsch, M. C Payne, G Csanyi, and A De Vita, *Low-speed fracture instabilities in a brittle crystal*, Nature, **455**, no. 7217, 1224, 2008.
13. S. Kohlhoff, P. Gumbsch, and H. F. Fischmeister, *Crack propagation in b.c.c. crystals studied with a combined finite-element and atomistic model*, Phil. Mag. A, **64**, 851–878, 1991.
14. E. B. Tadmor, R. Phillips, and M. Ortiz, *Mixed atomistic and continuum models of deformation in solids*, Langmuir, **12**, 4529–4534, 1996.
15. V. B. Shenoy, R. Miller, E. B. Tadmor, D. Rodney, R. Phillips, and M. Ortiz, *An adaptive finite element approach to atomic-scale mechanics — the quasicontinuum method*, J. Mech. Phys. Solids, **47**, 611–642, 1999.
16. V. B. Shenoy, R. Miller, E. B. Tadmor, R. Phillips, and M. Ortiz, *Quasicontinuum Models of Interfacial Structure and Deformation*, Phys. Rev. Lett., **80**, no. 4, 742–745, Jan 1998.
17. R. Miller, M. Ortiz, R. Phillips, V. B. Shenoy, and E. B. Tadmor, *Quasicontinuum models of fracture and plasticity*, Eng. Fract. Mech., **61**, 427–444, 1998.
18. D. R. Bowler, M. Fearn, C. M. Goringe, A. P. Horsfield, and D. G. Pettifor, *Hydrogen diffusion on Si(001) studied with the local density approximation and tight binding*, J. Phys.: Cond Mat, **10**, 3719, 1998.
19. Gábor Csányi, T. Albaret, G. Moras, M. C. Payne, and A. De Vita, *Multiscale hybrid simulation methods for material systems*, J. Phys.: Cond Mat, **17**, R691–R703, 2005.
20. Hai Lin and Donald G. Truhlar, *QM/MM: what have we learned, where are we, and where do we go from here?*, Theor. Chem. Acc., **117**, 185–199, 2007.
21. M. F. Ruiz-Lopez, *Combined QM/MM calculations in chemistry and biochemistry*, J. Mol. Struct.: THEOCHEM, **632**, ix, 2003.
22. Feliu Maseras and Keiji Morokuma, *IMOMM: A new integrated* ab initio + *molecu-*

lar mechanics geometry optimization scheme of equilibrium structures and transition states, J. Comput. Chem., **16**, 1170–1179, 1995.
23. M. Svensson, S. Humbel, R.D.J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, *ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt(P(t-Bu)$_3$)$_2$ + H_2 Oxidative Addition*, J. Phys. Chem., **100**, no. 50, 19357–19363, 1996.
24. Thom Vreven, Keiji Morokuma, Ödön Farkas, H. Bernhard Schlegel, and Michael J. Frisch, *Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints*, J. Comput. Chem., **24**, 760–769, 2003.
25. Alessandro Laio, Joost VandeVondele, and Ursula Rothlisberger, *A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations*, J. Chem. Phys., **116**, no. 16, 6941–6947, 2002.
26. N. Reuter, A. Dejaegere, B. Maigret, and M. Karplus, *Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches*, J. Phys. Chem. A, **104**, no. 8, 1720–1735, 2000.
27. J. C. H. Spence, Y. M. Huang, and O. Sankey, *Lattice trapping and surface reconstruction for silicon cleavage on (111). Ab-initio quantum molecular dynamics calculations*, Acta Metall. Mater., **41**, 2815–2824, 1993.
28. Jeremy Q. Broughton, Farid F. Abraham, Noam Bernstein, and Efthimios Kaxiras, *Concurrent coupling of length scales: Methodology and application*, Phys. Rev. B, **60**, no. 4, 2391–2403, Jul 1999.
29. Shuji Ogata, Fuyuki Shimojo, Rajiv K. Kalia, Aiichiro Nakano, and Priya Vashishta, *Hybrid quantum mechanical/molecular dynamics simulation on parallel computers: density functional theory on real-space multigrids*, Comput. Phys. Commun., **149**, 30–38, 2002.
30. Shuji Ogata, Elefterios Lidorikis, Fuyuki Shimojo, Aiichiro Nakano, Priya Vashishta, and Rajiv K. Kalia, *Hybrid finite-element/molecular-dynamics/electronic-density-functional approach to materials simulations on parallel computers*, Comput. Phys. Commun., **138**, 143–154, 2001.
31. Shuji Ogata, Fuyuki Shimojo, Rajiv K. Kalia, Aiichiro Nakano, and Priya Vashishta, *Environmental effects of H_2O on fracture initiation in silicon: A hybrid electronic-density-functional/molecular-dynamics study*, J. Appl. Phys., **95**, no. 10, 5316–5323, 2004.
32. Shuji Ogata, *Buffered-cluster method for hybridization of density-functional theory and classical molecular dynamics: Application to stress-dependent reaction of H_2O on nanostructured Si*, Phys. Rev. B., **72**, no. 4, 045348, 2005.
33. N. Bernstein, *Linear scaling nonorthogonal tight-binding molecular dynamics for nonperiodic systems*, Europhys. Lett., **55**, no. 1, 52–58, 2001.
34. N. Bernstein and D. W. Hess, *Lattice Trapping Barriers to Brittle Fracture*, Phys. Rev. Lett., **91**, 025501, 2003.

Bond-Order Potentials for Bridging the Electronic to Atomistic Modelling Hierarchies

Thomas Hammerschmidt and Ralf Drautz

Interdisciplinary Centre for Advanced Materials Simulation (ICAMS)
Ruhr-Universität Bochum
Stiepeler Strasse 129, 44801 Bochum, Germany
E-mail: {*thomas.hammerschmidt, ralf.drautz*}*@rub.de*

Robust interatomic potentials must be able to describe the making and breaking of interatomic bonds in a computationally efficient format so that the potentials may be employed in large-scale atomistic simulations. We summarize the fundamentals of such potentials, the bond-order potentials, and their derivation from the electronic structure. By coarse graining the tight-binding electronic structure and relating it to the local atomic environment, the bond-order potentials are derived as quantum-mechanical footed effective interatomic interactions.

1 What are Bond-Order Potentials?

Bond-order potentials are interatomic potentials that are derived from quantum mechanics. In contrast to classical empirical potentials, bond-order potentials capture bond formation and breaking, saturated and unsaturated bonds, dangling bonds and radical bonds, as well as single, double or triple bonds. The bond-order potentials provide similar accuracy as tight-binding calculations at less computational effort, and thus open the way to large-scale atomistic simulations of systems which cannot be described by classical empirical potentials.

The bond-order potentials (BOPs) are derived by systematically coarse graining the electronic structure at two levels of approximation,

1. In the first step, the density functional theory (DFT) formalism is approximated in terms of physically and chemically intuitive contributions within the tight-binding (TB) bond model[1,2]. The TB approximation is sufficiently accurate to predict structural trends across the sp-valent and d-valent elements, as well as sufficiently simple to allow a physically meaningful interpretation of the bonding in terms of σ, π and δ contributions. The parameters of the TB model can be obtained from ab-initio calculations in a systematic way.

2. In the second step, the TB electronic structure is coarse grained and related to the local topology and coordination of the material. The functional form of the bond energy is derived as a function of the positions and the types of atoms that surround a given bond.

The first step of coarse graining from DFT to TB is discussed by Anthony Paxton in his contribution to this issue[1]. In this contribution we will start from the TB description of the electronic structure, focus on the second level of coarse graining and discuss how the electronic structure may be hidden in effective interatomic interactions.

1.1 Aim of this Contribution and some Literature

With this short introduction we do not aim at giving an overview of the bond-order potentials. Instead, with our contribution we would like to give an easy to read summary of the ideas and concepts that are used to coarse grain the electronic structure into effective interatomic interactions. A recent special issue *Modelling Electrons and Atoms for Materials Science* of Progress in Materials Science[3] contains a number of reviews that give a detailed overview of the bond-order potentials and their application to the simulation of elastic and plastic properties of transition metals, the growth of semiconductor thin films and hydrocarbons[4–7]. We would like to recommend these reviews to the interested reader.

2 Binding Energy

The starting point for the derivation of bond-order potentials is the tight-binding bond model[2] that is introduced in the lecture of Anthony Paxton[1]. The binding energy within the tight-binding bond model is given as the sum of covalent bond energy U_{bond}, promotion energy U_{prom}, and repulsive energy U_{rep} ,

$$U_B = U_{\mathrm{bond}} + U_{\mathrm{prom}} + U_{\mathrm{rep}} \,. \tag{1}$$

The promotion energy is calculated as a sum over orbitals $|i\alpha\rangle$ centred on atom i (where α labels the valence orbital), whereas the repulsive energy is often approximated as sum over pairs of atoms

$$U_{\mathrm{prom}} = \sum_{i\alpha} E^{(0)}_{i\alpha} \left(N_{i\alpha} - N^{(0)}_{i\alpha} \right) \,, \tag{2}$$

$$U_{\mathrm{rep}} = \sum_{ij} \phi_{ij} \left(R_{ij} \right) \,, \tag{3}$$

with the free atom reference onsite levels $E^{(0)}_{i\alpha}$. The promotion energy accounts for the redistribution of the electrons across the orbitals of an atom due to hybridisation. The simplest form of the repulsive energy as given above is a pairwise term that depends solely on the interatomic distance R_{ij} between atoms i and j. Some materials require a more complex description of the repulsive energy, *e.g.* Mrovec *et al.*[8] introduced a Yukawa-type environment-dependent term to account for the strong core repulsion in transition metals.

As we will see in the following, the bond energy U_{bond} can be given in either onsite representation (in terms of the atom-based density of states) or intersite representation (in terms of the bond-based density matrix or bond order). The two representations are equivalent but offer different views on the formation of bonds in materials.

2.1 Bond Energy: Onsite Representation

The onsite representation of the bond energy is based on the local density of states $n_{i\alpha}(E)$ of orbital α on atom i. The contribution of each orbital to the bond energy is calculated by integrating its local density of states (DOS) up to the Fermi level E_{F}

$$U_{\mathrm{bond}} = 2 \sum_{i\alpha} \int^{E_{\mathrm{F}}} \left(E - E_{i\alpha} \right) n_{i\alpha}(E) \mathrm{d}E \,. \tag{4}$$

The factor of two accounts for the neglect of magnetism in the model, so that spin up and spin down spin channels are degenerate. The onsite level $E_{i\alpha}$ is shifted relative to its free atom value $E_{i\alpha}^{(0)}$ until self-consistency is achieved (*cf.* lecture of Anthony Paxton[1]). The local density of states $n_{i\alpha}(E)$ may be obtained from the eigenfunctions $|\psi_n\rangle$ of the Hamiltonian $\hat{H}$,

$$\hat{H}\psi_n = E_n\psi_n\,, \tag{5}$$

by expressing the eigenfunctions $|\psi_n\rangle$ in an orthonormal basis centred on atoms i

$$|\psi_n\rangle = \sum_{i\alpha} c_{i\alpha}^{(n)}|i\alpha\rangle\,, \tag{6}$$

where the index α denotes the valence orbital. Then, by calculating the band energy $U_{\rm band}$ as the sum over all occupied orbitals, we find that

$$\begin{aligned} U_{\rm band} &= 2\sum_{n}^{\rm occ} E_n = 2\sum_{n}^{\rm occ}\langle\psi_n|\hat{H}|\psi_n\rangle = 2\sum_{n}^{\rm occ} E_n\langle\psi_n|\psi_n\rangle \\ &= 2\sum_{n}^{\rm occ}\sum_{i\alpha}\sum_{j\beta} E_n c_{i\alpha}^{*(n)} c_{j\beta}^{(n)}\langle i\alpha|j\beta\rangle = 2\sum_{n}^{\rm occ}\sum_{i\alpha}\sum_{j\beta} E_n c_{i\alpha}^{*(n)} c_{j\beta}^{(n)}\delta_{i\alpha j\beta} \\ &= 2\sum_{i\alpha}\int^{E_{\rm F}} E\sum_{n}\delta(E-E_n)c_{i\alpha}^{*(n)}c_{i\alpha}^{(n)}{\rm d}E\,. \end{aligned} \tag{7}$$

We identify the local density of states of orbital $|i\alpha\rangle$ as

$$n_{i\alpha} = \sum_{n}\left|c_{i\alpha}^{(n)}\right|^2\delta(E-E_n)\,, \tag{8}$$

such that the band energy $U_{\rm band}$ is written as

$$U_{\rm band} = 2\sum_{i\alpha}\int^{E_{\rm F}} E n_{i\alpha}(E){\rm d}E\,. \tag{9}$$

The bond energy Eq.(4) is the band energy calculated with respect to the onsite levels $E_{i\alpha}$,

$$U_{\rm bond} = 2\sum_{i\alpha}\int^{E_{\rm F}}(E-E_{i\alpha})\,n_{i\alpha}(E){\rm d}E = U_{\rm band} - \sum_{i\alpha}E_{i\alpha}N_{i\alpha}\,, \tag{10}$$

with the number of electrons $N_{i\alpha}$ in orbital $|i\alpha\rangle$,

$$N_{i\alpha} = 2\sum_{i\alpha}\int^{E_{\rm F}} n_{i\alpha}(E){\rm d}E\,. \tag{11}$$

Some authors prefer not to calculate the *bond* energy that is calculated with respect to the onsite levels but to use the *band* energy instead, such that

$$U_B = U_{\rm band} + U_{\rm prom} + U_{\rm rep}\,. \tag{12}$$

However, as discussed in the lecture of Anthony Paxton[1], this tight-binding *band* model is inconsistent with the force theorem[2,9,10] while the bond energy in the tight-binding *bond* model properly accounts for the redistribution of charge due to the shift of the onsite levels that arise from atomic displacements.

2.2 Bond Energy: Intersite Representation

An alternative but equivalent representation to the onsite representation of the band energy Eq.(9) is the intersite representation. The intersite representation is obtained by expanding the eigenfunctions $|\psi_n\rangle = \sum_{i\alpha} c_{i\alpha}^{(n)} |i\alpha\rangle$ in terms of the TB basis,

$$\begin{aligned} U_{\mathrm{band}} &= 2\sum_{n}^{\mathrm{occ}} E_n = 2\sum_{n}^{\mathrm{occ}} \langle\psi_n|\hat{H}|\psi_n\rangle = 2\sum_{i\alpha}\sum_{j\beta}\sum_{n}^{\mathrm{occ}} c_{i\alpha}^{*(n)} c_{j\beta}^{(n)} \langle i\alpha|\hat{H}|j\beta\rangle \\ &= 2\sum_{i\alpha j\beta} \rho_{i\alpha j\beta} H_{i\alpha j\beta}\,, \end{aligned} \tag{13}$$

with the density matrix

$$\rho_{i\alpha j\beta} = \sum_{n}^{\mathrm{occ}} c_{i\alpha}^{*(n)} c_{j\beta}^{(n)}\,. \tag{14}$$

The *bond* energy is obtained from the *band* energy in intersite representation by restricting the summation to off-diagonal elements as $N_{i\alpha} = \rho_{i\alpha i\alpha}$. Therefore, the bond energy in intersite representation is given by

$$U_{\mathrm{bond}} = 2\sum_{i\alpha\neq j\beta} \rho_{i\alpha j\beta} H_{i\alpha j\beta}\,. \tag{15}$$

The bond order $\Theta_{i\alpha j\beta}$ of a bond between the valence orbitals α and β of two atoms i and j is just two times the corresponding element of the density matrix

$$\Theta_{i\alpha j\beta} = 2\rho_{i\alpha j\beta}\,. \tag{16}$$

By construction the onsite and intersite representation of the bond energy are equivalent

$$U_{\mathrm{bond}} = 2\sum_{i\alpha}\int^{E_{\mathrm{F}}} (E - E_{i\alpha})\, n_{i\alpha}(E)\mathrm{d}E = \sum_{i\alpha\neq j\beta} \Theta_{i\alpha j\beta} H_{i\alpha j\beta}\,, \tag{17}$$

however, the two representations offer different views on bond formation. We see that while the bond energy in onsite representation is obtained by filling electrons into the local density of states $n_{i\alpha}(E)$ on each atom, the intersite representation calculates the bond energy as a sum over pairwise Hamiltonian matrix elements $H_{i\alpha j\beta}$ that are weighted with the density matrix element $\rho_{i\alpha j\beta}$. In the following we will discuss some properties of the density matrix.

3 Properties of the Bond Order

In the previous section we decomposed *global* quantities, like the bond energy or the band energy, in their *local* contributions, the atom-based local density of states in the onsite representation and the bond-based bond order in the intersite representation. In the following we will discuss some properties of the bond order, while the properties of the local density of states will be discussed in section 4.

An intuitive physical interpretation of the bond order becomes apparent when we transform the atomic orbitals to linear combinations (dimer orbitals)

$$|+\rangle = \frac{1}{\sqrt{2}}\left(|i\alpha\rangle + |j\beta\rangle\right) \qquad \text{bonding}\,, \tag{18}$$

$$|-\rangle = \frac{1}{\sqrt{2}}\left(|i\alpha\rangle - |j\beta\rangle\right) \qquad \text{antibonding}\,. \tag{19}$$

The number of electrons in the bonding and antibonding dimer orbitals may be obtained by projection on the occupied eigenstates,

$$N_+ = 2\sum_n^{\text{occ}} |\langle +|\psi_n\rangle|^2\,, \quad N_- = 2\sum_n^{\text{occ}} |\langle -|\psi_n\rangle|^2\,, \tag{20}$$

By expanding the eigenstates in the atomic basis, Eq.(6), and by making use of the definition of the bond order Eq.(16), one finds that the bond order is one-half the difference between the number of electrons in the bonding state compared to the antibonding state

$$\Theta_{i\alpha j\beta} = \frac{1}{2}\left(N_+ - N_-\right)\,. \tag{21}$$

With a maximum of two electrons in an orbital, the bond order takes its largest absolute value of 1 for two electrons of opposite spin in the bonding and and none in the antibonding orbital. Furthermore, as the number of electrons in the bonding state N_+ is less or equal to the total number of electrons in the bond $N_{i\alpha j\beta} = \frac{1}{2}\left(N_{i\alpha} + N_{i\beta}\right) = \frac{1}{2}\left(N_+ + N_-\right)$,

$$N_+ \leq N_+ + N_-\,, \tag{22}$$

the value of the bond order in general is limited by an envelope function[11]

$$|\Theta_{i\alpha j\beta}| \leq \begin{cases} N_{i\alpha j\beta} & \text{for } 0 \leq N_{i\alpha j\beta} \leq 1\,, \\ 2 - N_{i\alpha j\beta} & \text{for } 1 < N_{i\alpha j\beta} \leq 2\,. \end{cases} \tag{23}$$

As an example, consider the H_2 molecule with one s-orbital on each atom. The eigenstates of the H_2 dimer are given by bonding and antibonding linear combinations of the s-orbitals. Both valence electrons occupy the bonding state, while the antibonding state remains empty. Therefore we expect that the H_2 dimer forms a fully saturated covalent bond with bond order $\Theta = 1$. If we look at a He_2 dimer instead, the eigenstates are also given by bonding and antibonding linear combinations of the s-orbitals just like in the case of H_2. However, now the 4 valence electrons have to completely fill both, the bonding and the antibonding states such that the bond order is zero $\Theta = 0$. Therefore, we expect that the He_2 molecule does not form a covalent bond. In contrast to these two extremal cases, the bond order usually takes intermediate values (see Fig. 1) that depend sensitively on local coordination and number of valence electrons. It is the aim of the bond-order potentials to describe these intermediate values as accurately as possible. More examples and a detailed discussion of the bond order of different molecules and solids is given in the textbook of Pettifor[12].

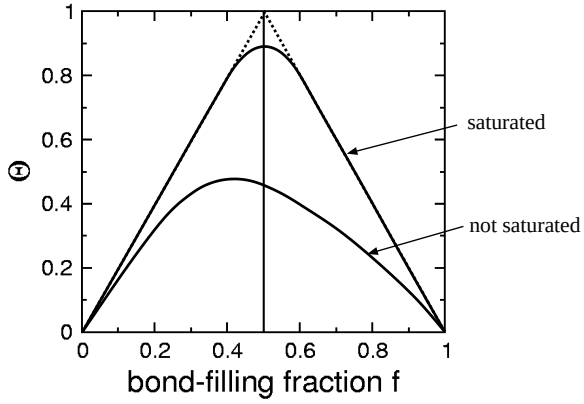


Figure 1. Schematic of the bond order as a function of the number of electrons in the bond (bond-filling fraction $f = N_{i\alpha j\beta}/2$). The bond order of a saturated bond closely follows the envelope function Eq.(23) and is close to 1 at half-full band. Typically materials with open structures like, for example, Si in the diamond lattice, show the formation of saturated bonds. In close-packed crystals, for example in d-valent transition metals, the electrons cannot be distributed only into bonding states, the bonds are not saturated and the bond order takes a smaller value.

4 Moments

For the development of effective interatomic potentials we would like to bypass the numerical diagonalisation of the TB Hamiltonian $\hat{H}$ and instead determine local quantities like the local density of states $n_{i\alpha}(E)$ or the bond order $\Theta_{i\alpha j\beta}$ directly from the local atomic environment. This may be achieved by making use of the moments theorem[13] that relates the electronic structure ($n_{i\alpha}(E)$, $\rho_{i\alpha j\beta}$) to the crystal structure (the position of the atoms). The Nth moment of orbital $|i\alpha\rangle$ is given by

$$\mu_{i\alpha}^{(N)} = \int E^N n_{i\alpha}(E) \mathrm{d}E \,. \tag{24}$$

Inserting the density of states from Eq.(8) and making use of the identity operator

$$\hat{1} = \sum_n |\psi_n\rangle\langle\psi_n| \,, \tag{25}$$

results in an expression for the moments in terms of atomic orbitals $|i\alpha\rangle$ and the Hamiltonian $\hat{H}$:

$$\begin{aligned}
\mu_{i\alpha}^{(N)} &= \int E^N \sum_n \left| c_{i\alpha}^{(n)} \right|^2 \delta(E - E_n) \mathrm{d}E \\
&= \sum_n \left| \langle i\alpha | \psi_n \rangle \right|^2 E_n^N \\
&= \sum_n \langle i\alpha | \hat{H}^N | \psi_n \rangle \langle \psi_n | i\alpha \rangle \\
&= \langle i\alpha | \hat{H}^N | i\alpha \rangle \,.
\end{aligned} \tag{26}$$

By using an orthogonal basis set that completely spans the TB Hilbert space, *i.e.*

$$\hat{1} = \sum_{j\beta} |j\beta\rangle\langle j\beta| \,, \tag{27}$$

the Nth power of the Hamilton operator acting on orbital $|i\alpha\rangle$ can be written as the product of N Hamilton matrices,

$$\langle i\alpha|\hat{H}^N|i\alpha\rangle = \sum_{j\beta k\gamma\cdots} \langle i\alpha|\hat{H}|j\beta\rangle\langle j\beta|\hat{H}|k\gamma\rangle\langle k\gamma|\hat{H}|\dots\rangle\cdots\langle\dots|\hat{H}|i\alpha\rangle \,. \tag{28}$$

Each Hamiltonian matrix element $H_{i\alpha j\beta} = \langle i\alpha|\hat{H}|j\beta\rangle$ connects two neighbouring atoms i and j and is frequently called a *hop*. Looking at the indices, we see that the product of Hamiltonian matrices defines a path through the atomic structure ($|i\alpha\rangle \to |j\beta\rangle \to |k\gamma\rangle \to \cdots \to |i\alpha\rangle$) which we will refer to as hopping path. Therefore the Nth moment $\mu_{i\alpha}^{(N)}$, Eq.(26), can be understood as the sum over all hopping paths of length N that start and end on the same orbital $|i\alpha\rangle$,

$$\mu_{i\alpha}^{(N)} = \int E^N n_{i\alpha}(E)\mathrm{d}E = \sum_{j\beta k\gamma\cdots} H_{i\alpha j\beta}H_{j\beta k\gamma}H_{k\gamma\dots}\cdots H_{\dots i\alpha} \,. \tag{29}$$

Figure 2 illustrates one hopping path that contributes to the 4th moment. As different crystal structures have different numbers of hopping paths of a given lengths, the moments are sensitive to changes in the crystal structure. Higher moments correspond to longer hopping paths and thus to a more far-sighted sampling of the atomic environment.

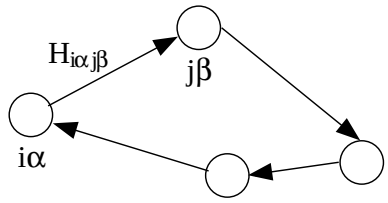


Figure 2. A path that contributes to the 4th moment of orbital $i\alpha$. The 4th moment is important for the energy difference of the *fcc* and *bcc* structure of the transition metals.

Moments are well known in statistical mathematics as a concept to describe a distribution (in our case the local DOS). The first few moments are often discussed as measures of specific properties of the distribution,

$$\mu_{i\alpha}^{(0)} = \int n_{i\alpha}(E)\mathrm{d}E \quad : \mathrm{norm} \,, \tag{30}$$

$$\mu_{i\alpha}^{(1)} = \int E n_{i\alpha}(E)\mathrm{d}E \ : \mathrm{centre\,of\,gravity} \,, \tag{31}$$

$$\mu_{i\alpha}^{(2)} = \int E^2 n_{i\alpha}(E)\mathrm{d}E : \mathrm{root\,mean\,square\,width} \,, \tag{32}$$

$$\mu_{i\alpha}^{(3)} = \int E^3 n_{i\alpha}(E)\mathrm{d}E : \mathrm{skewness} \,, \tag{33}$$

$$\mu_{i\alpha}^{(4)} = \int E^4 n_{i\alpha}(E)\mathrm{d}E : \mathrm{bimodality} \,. \tag{34}$$

While $\mu_{i\alpha}^{(0)}$ and $\mu_{i\alpha}^{(1)}$ do not contain any information of the surroundings of an atom, the second moment $\mu_{i\alpha}^{(2)}$ is the lowest moment that contains physical information of the environment of an atom (the root mean square width of the density of states). The Hamiltonian is typically attractive, therefore the third moment is typically negative

$$\mu_{i\alpha}^{(3)} = \sum_{j\beta k\gamma} \langle i\alpha|\hat{H}|j\beta\rangle\langle j\beta|\hat{H}|k\gamma\rangle\langle k\gamma|\hat{H}|i\alpha\rangle < 0\,, \tag{35}$$

and gives rise to a skewed DOS as illustrated in Fig. 3(a). Therefore, if one calculates the energy difference of two densities of states at identical second moment but with $\mu_{i\alpha}^{(3)} = 0$ and $\mu_{i\alpha}^{(3)} < 0$, one obtains as a function of band filling a figure similar to Fig. 3(b). For less than half-full band the negative 3rd moment contribution tends to stabilise the DOS with $\mu_{i\alpha}^{(3)} < 0$ relative to the DOS with vanishing third moment. The third moment gives a first

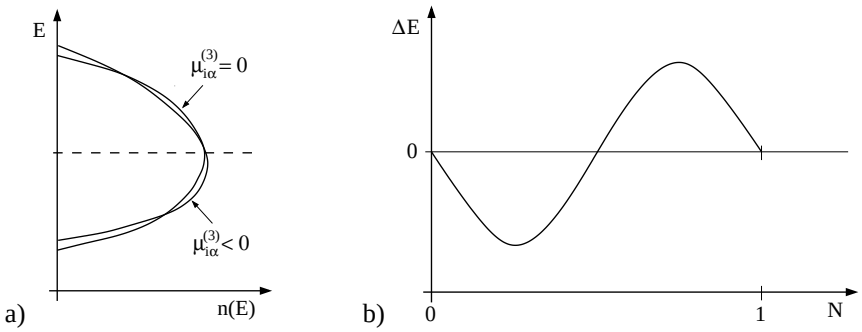


Figure 3. The 3rd moment gives rise to a skewing of the DOS (a) that typically (for $\mu_{i\alpha}^{(3)} < 0$) stabilises close-packed structures for less than half-full band (b).

indication of the crystal structure of elements with less than half-full band (like Mg and Al): the observed close-packed structure offers many self-returning paths of length three and therefore has a large third moment. In contrast, elements with more than half-full band (like Cl and S) tend to avoid a large third moment and therefore form open structures or molecules that have no hopping paths of length three.

The fourth moment characterises the bi-modal (in contrast to uni-modal) behaviour of the density of states as shown in Fig. 4(a). A bimodal DOS has a low density of states at the centre of the band and tends to be stable over a unimodal DOS at half-full band as shown in Fig. 4(b). This is the reason why sp-valent elements with half-full band (such as Si, Ge) have a tendency to crystallise in the diamond structure. The discussion of the first four moments may be generalised for higher moments. For example, six moments are required to resolve the energy difference between the close-packed *fcc* and *hcp* lattices[14], many of the small differences between more complex crystal structures can also be resolved with an expansion to only about the 6th moment[15,16]. Furthermore, if two structures are different only at the level of the Nth moment and this Nth moment dominates, then the energy difference between the two structures shows $N-2$ zeros between empty and full band[17].

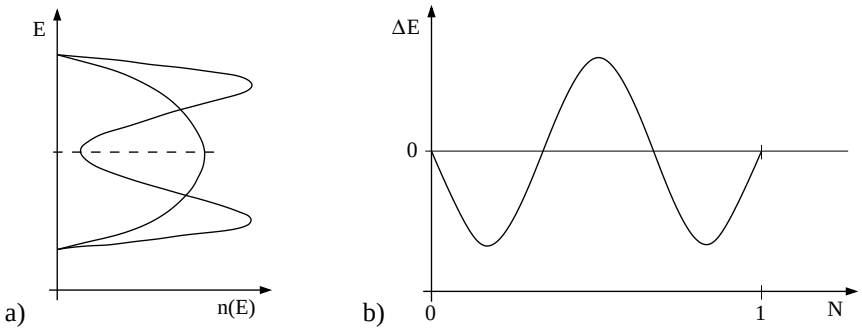


Figure 4. The 4th moment causes the DOS to take a bimodal shape (a), thereby favouring the diamond structure at half-full band.

5 Recursion

In the previous section we showed that the moments of the density of states relate the atomic structure to the electronic structure. A mathematically equivalent way of relating the electronic structure to the crystal structure is the recursion method[18].
Given a starting state $|u_0\rangle$, which we may think of for example as an atomic orbital $|i\alpha\rangle$, the Hamilton operator is used to generate a new state $|u_1\rangle$ by

$$b_1|u_1\rangle = \left(\hat{H} - a_0\right)|u_0\rangle\,. \tag{36}$$

The new state is normalized ($\langle u_1|u_1\rangle = 1$) and orthogonal to $|u_0\rangle$ ($\langle u_1|u_0\rangle = 0$). The coefficients a_0 and b_1 are determined by multiplying from the left with $|u_1\rangle$ and $|u_0\rangle$:

$$b_1 = \langle u_1|\hat{H}|u_0\rangle\,, \tag{37}$$
$$a_0 = \langle u_0|\hat{H}|u_0\rangle\,. \tag{38}$$

In a similar fashion, the Hamiltonian is used to generate from $|u_1\rangle$ an other new state $|u_2\rangle$ that cannot be written as a linear combination of $|u_0\rangle$ and $|u_1\rangle$:

$$b_2|u_2\rangle = \left(\hat{H} - a_1\right)|u_1\rangle - b_1|u_0\rangle\,, \tag{39}$$

which is again normalized ($\langle u_2|u_2\rangle = 1$) and orthogonal to $|u_1\rangle$ ($\langle u_2|u_1\rangle = 0$). The coefficients a_1 and b_2 are given correspondingly by

$$b_2 = \langle u_2|\hat{H}|u_1\rangle\,, \tag{40}$$
$$a_1 = \langle u_1|\hat{H}|u_1\rangle\,. \tag{41}$$

The general form of the recursion may be written as

$$b_{n+1}|u_{n+1}\rangle = \left(\hat{H} - a_n\right)|u_n\rangle - b_n|u_{n-1}\rangle\,, \tag{42}$$

with the matrix elements

$$b_n = \langle u_n|\hat{H}|u_{n-1}\rangle\,, \tag{43}$$
$$a_n = \langle u_n|\hat{H}|u_n\rangle\,. \tag{44}$$

The states $|u_n\rangle$ are orthogonal, $\langle u_n|u_m\rangle = \delta_{nm}$. This means that in the basis $\{|u_0\rangle, |u_1\rangle, |u_2\rangle, \dots\}$, which is generated from the atomic-like orbitals $|u_0\rangle = |i\alpha\rangle$ by recursion, the Hamiltonian matrix takes the following, tridiagonal form

$$\langle u_n|\hat{H}|u_m\rangle = \begin{pmatrix} a_0 & b_1 & & & & & \\ b_1 & a_1 & b_2 & & & & \\ & b_2 & a_2 & b_3 & & & \\ & & b_3 & a_3 & b_4 & & \\ & & & b_4 & a_4 & \ddots & \\ & & & & \ddots & \ddots & \ddots \\ & & & & & \ddots & \ddots \end{pmatrix} .$$

All elements that are not in the diagonal or next to the diagonal are identical to zero. This Hamiltonian matrix may be thought of as the Hamiltonian of a one-dimensional chain with only nearest neighbour hopping matrix elements, see Fig. 5. Using the recursion

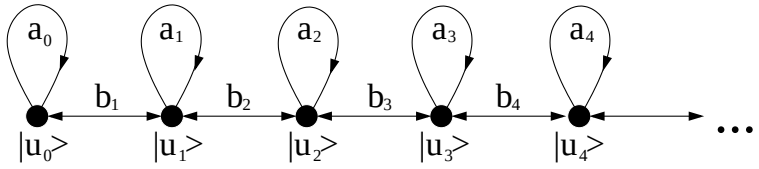


Figure 5. Graphical representation of the recursion Hamiltonian as one-dimensional chain: the Lanczos chain.

and writing $|u_n\rangle$ as linear combination of atomic orbitals, the moments are related to the expansion coefficients a_n and b_n. The Nth moment can be determined by summing over all possible paths of length N that start and end on orbital $|u_0\rangle$. For example, the first four moments are given by

$$\mu_{i\alpha}^{(0)} = 1\,, \tag{45}$$

$$\mu_{i\alpha}^{(1)} = a_0\,, \tag{46}$$

$$\mu_{i\alpha}^{(2)} = a_0^2 + b_1^2\,, \tag{47}$$

$$\mu_{i\alpha}^{(3)} = a_0^3 + 2a_0b_1^2 + a_1b_1^2\,, \tag{48}$$

which is easily verified by identifying all paths of corresponding length in Fig. 5. The purpose of introducing the recursion method in the context of bond-order potentials is to transform the TB Hamiltonian to an orthogonal basis where it takes a tridiagonal form. This procedure of transforming the Hamiltonian to a semi-infinite one-dimensional nearest-neighbour chain is the Lanczos algorithm[19] and establishes an $\mathcal{O}$(N) approach to calculate the local electronic density of states as we shall see in the following.

6 Green's Functions

In the previous section we learned how to calculate the moments of the density of states from the crystal structure. We would like to use the information contained in the moments

to calculate the bond energy U_{bond}, Eq.(4). The Green's function $\hat{G}$ is closely related to the density of states Eq.(8) and the density matrix Eq.(14), as we will see in the following. It will therefore be helpful for the construction of the bond energy U_{bond}, Eqs.(4) and (15). As a first step in this direction we will use the Green's functions to reconstruct the local density of states $n_{i\alpha}(E)$ from the moments. Once we have obtained the local density of states, we can integrate it to calculate the bond energy. We define the Green's function $\hat{G}$ as the inverse of the Hamiltonian,

$$\hat{G} = \left(E\hat{1} - \hat{H}\right)^{-1} . \tag{49}$$

As the Hamilton operator in the basis of the eigenstates ψ_n is written as

$$\langle\psi_n| \left(E\hat{1} - \hat{H}\right) |\psi_m\rangle = (E_n - E)\delta_{nm} , \tag{50}$$

and by definition of $\hat{G}$,

$$\langle\psi_n| \left(E\hat{1} - \hat{H}\right) \hat{G}|\psi_m\rangle = \langle\psi_n|\psi_m\rangle = \delta_{nm} , \tag{51}$$

the Green's function matrix elements of the eigenstates may be written explicitly as

$$\langle\psi_n|\hat{G}|\psi_m\rangle = \frac{\delta_{nm}}{E - E_n} . \tag{52}$$

This can be verified by inserting the identity $\hat{1} = \sum_k |\psi_k\rangle\langle\psi_k|$,

$$\langle\psi_n| \left(E\hat{1} - \hat{H}\right) \hat{G}|\psi_m\rangle = \sum_k \langle\psi_n| \left(E\hat{1} - \hat{H}\right) |\psi_k\rangle\langle\psi_k|\hat{G}|\psi_m\rangle \tag{53}$$

$$= \sum_k (E - E_n)\, \delta_{nk} \frac{\delta_{km}}{E - E_m} \tag{54}$$

$$= (E - E_n)\, \delta_{nm} \frac{1}{E - E_m} \tag{55}$$

$$= \delta_{nm} . \tag{56}$$

The matrix elements of the Green's function in the atomic orbital basis $G_{i\alpha j\beta}(E) = \langle i\alpha|\hat{G}|j\beta\rangle$ are obtained as

$$G_{i\alpha j\beta}(E) = \sum_{nm} \langle i\alpha|\psi_n\rangle\langle\psi_n|\hat{G}|\psi_m\rangle\langle\psi_m|j\beta\rangle = \sum_n \frac{c_{i\alpha}^{*(n)} c_{j\beta}^{(n)}}{E - E_n} , \tag{57}$$

By making use of the identity/residue

$$-\frac{1}{\pi}\mathrm{Im} \int \frac{1}{E - E_n}\mathrm{d}E = \int \delta(E - E_n)\mathrm{d}E , \tag{58}$$

we can replace Eq.(8) and Eq.(14) using matrix elements of the Green's function

$$n_{i\alpha}(E) = -\frac{1}{\pi}\mathrm{Im}\, G_{i\alpha i\alpha}(E) , \tag{59}$$

$$\rho_{i\alpha j\beta} = -\frac{1}{\pi}\mathrm{Im} \int^{E_\mathrm{F}} G_{i\alpha j\beta}\mathrm{d}E . \tag{60}$$

Connection can now be made to the recursion method introduced in the previous section. The diagonal element of the Green's function at the starting orbital of the semi-infinite one-dimensional Lanczos chain is given as a continued fraction[20]

$$G_{i\alpha i\alpha} = G_{00} = \cfrac{1}{E - a_0 - \cfrac{b_1^2}{E - a_1 - \cfrac{b_2^2}{E - a_2 - \cfrac{b_3^2}{\ddots}}}}. \tag{61}$$

The continued fraction expansion provides a direct way of calculating the density of states which in turn may be used to calculate the bond energy.
Taking the continued fraction to an infinite number of recursion levels corresponds to an exact solution of the tight-binding model. By *terminating* the continued fraction after a certain number of levels, a local expansion of the electronic structure is obtained. The different flavors of using truncated Green's function expansion for a local calculation of the bond energy are presented in the following section. A more detailed review of the connection between bond-order potentials, Green's functions and the recursion method is given, *e.g.*, in Refs. 21–23.

7 Calculation of the Bond-Energy I – Numerical Bond-Order Potentials

The recursion expansion representation of the Hamiltonian Eq. (42) offers a direct way of writing the onsite Greens-function matrix elements $G_{i\alpha i\alpha} = \langle i\alpha|\hat{G}|i\alpha\rangle = G_{00}$ in the form of a continued fraction expansion, Eq.(61). For the bond-order potentials we are interested in a local calculation of the bond energy and not in an exact solution of the underlying TB model. This is achieved by *terminating* the expansion after a few recursion levels n. This is equivalent to evaluating the first $2n + 1$ moments of the density of states (cf. Sec. 5). In the simplest case, the recursion coefficients a_m and b_m for $m > n$ are replaced by a constant terminator

$$a_m = a_\infty, \quad b_m = b_\infty \qquad \text{for } m > n\,. \tag{62}$$

By inserting the continued fraction expression for the Green's function matrix element Eq.(61) in Eq.(59) one obtains an approximate closed-form representation of the density of states $n_{i\alpha}$. The bond energy Eq.(4) is obtained by *numerical* integration of

$$U_{\text{bond}} = -\frac{2}{\pi}\text{Im}\int^{E_\text{F}} (E - E_{i\alpha}) \cfrac{1}{E - a_0 - \cfrac{b_1^2}{\ddots - \cfrac{\ddots}{E - a_\infty - \cfrac{b_\infty^2}{E - \ddots}}}} \text{d}E \tag{63}$$

and therefore this representation of the energy is called numerical bond-order potential. In general the approximation error in the bond energy will become smaller with more

recursion levels n taken into account exactly. Therefore, the number of recursion levels provides a way of systematically converging the bond energy to the bond energy that one obtains from an exact solution of the TB Hamiltonian.
For the calculation of the forces on the atoms one requires the bond-order/density matrix and therefore the calculation of $G_{i\alpha j\beta}$. For numerical stability and convergence of the continued fraction expansion of $G_{i\alpha j\beta}$, one relates $G_{i\alpha j\beta}$ to the onsite matrix elements $G_{i\alpha i\alpha}$ and $G_{j\beta j\beta}$. This is achieved by using a linear combination of atomic orbitals in the recursion expansion

$$|u_0\rangle = \frac{1}{\sqrt{2}} \left(|i\alpha\rangle + \mathrm{e}^{i\vartheta} |j\beta\rangle \right) , \tag{64}$$

with $\vartheta = \cos(\lambda)$ such that

$$G_{00} = \lambda G_{i\alpha j\beta} + \frac{1}{2} \left(G_{i\alpha i\alpha} + G_{j\beta j\beta} \right) . \tag{65}$$

Therefore, the intersite matrix elements of the Green's function is given as a derivative of the onsite elements of the starting Lanczos orbital, a central result of BOP theory[24]

$$G_{i\alpha j\beta} = \frac{\mathrm{d}}{\mathrm{d}\lambda} G_{00} \Big|_{\lambda=0} . \tag{66}$$

At any level of approximation exists a termination of the expansion of $G_{i\alpha j\beta}$ which ensures that the onsite and intersite representation of the bond energy are identical[25], as of course it would have to be if the problem would have been solved exactly. A detailed review of the numerical bond-order potentials is available in Ref. 5.

8 Calculation of the Bond-Energy II – Analytic Bond-Order Potentials

As the integral for the calculation of the bond energy in Eq.(63) is carried out numerically in numerical BOPs, no analytic representation of the effective interactions between atoms and therefore no analytic interatomic potential may be obtained. In this section we will discuss how analytic representations of the bond energy may be obtained, such that explicit analytic interatomic potentials may be written down.

8.1 Analytic Bond-Order Potentials for Semiconductors

If the expansion of $G_{i\alpha i\alpha}$ in Eq.(61) is terminated with $a_\infty = 0$ and $b_\infty = 0$ after only two recursion levels ($n = 2$) corresponding to four moments, the integral for the bond energy Eq.(63) may be carried out analytically. In order to achieve a good convergence with only two recursion levels, the starting state of the recursion $|u_0\rangle$ must be taken as an as close approximation of the solution as possible. For semiconductors with saturated covalent bonds one achieves very good convergence if the starting state is chosen as a dimer orbital[26,27]

$$|u_0\rangle = \frac{1}{\sqrt{2}} \left(|i\alpha\rangle + |j\beta\rangle \right) . \tag{67}$$

The resulting analytic bond-order potentials[26,27] have been applied to modelling the growth of semiconductor films and hydrocarbons. A detailed review of the analytic BOPs for semiconductors may be found in Refs. 6,7. If one takes the expansion of this analytic bond-order potential only to two moments of the density of states instead of four moments, then an expansion is obtained that is very close[28] to the empirical potential given by Tersoff[29]. Therefore the analytic BOP may be viewed as an systematic extension of the Tersoff-Brenner-type potentials.

8.2 Analytic Bond-Order Potentials for Transition Metals

In a close-packed transition metal, the bonds between atoms are not saturated. Therefore the expansion of the analytic BOPs for semiconductors that is built on a saturated dimer bond may not be directly applied to transition metals. Instead of taking a dimer orbital as the starting state of the expansion, inserting a spherical atomic orbital into a close-packed crystal structure leads to a faster convergence of the expansion. However, in order to resolve for example the energy difference between the *fcc* and *hcp* structure in a canonical TB model[30], at least six moments are required[14]. For six moments or equivalently three recursion levels, the integration of Eq.(63) cannot be carried out analytically. Instead of integrating Eq.(63), one therefore constructs a perturbation expansion of the continued fraction representation of $G_{i\alpha i\alpha}$. This perturbation expansion may then be integrated analytically.

The starting point of the expansion is the observation that the Green's function may be written down in a compact form if all the expansion coefficients a_n and b_n are taken identical to

$$a_n = a_\infty \quad , \tag{68}$$

$$b_n = b_\infty \quad . \tag{69}$$

Then the density of states is given by

$$n_{i\alpha}^{(0)}(\varepsilon) = \frac{2}{\pi}\sqrt{1-\varepsilon^2}\,, \tag{70}$$

with the normalized energy ε,

$$\varepsilon = \frac{E - a_\infty}{2b_\infty} \quad . \tag{71}$$

The density of states $n_{i\alpha}^{(0)}(\varepsilon)$ is then used as the reference density of states in a perturbation expansion[31]

$$n_{i\alpha}(\varepsilon) = n_{i\alpha}^{(0)}(\varepsilon) + \delta n_{i\alpha}(\varepsilon)\,. \tag{72}$$

Chebyshev polynomials $P_n(\varepsilon)$ of the second kind are orthogonal with respect to the weight function $n_{i\alpha}^{(0)}$,

$$\frac{2}{\pi}\int_{-1}^{+1} P_n(\varepsilon)P_m(\varepsilon)\sqrt{1-\varepsilon^2}\mathrm{d}\varepsilon = \delta_{nm}\,. \tag{73}$$

The density of states is thus expanded in terms of Chebyshev polynomials

$$n_{i\alpha}(\varepsilon) = \frac{2}{\pi}\sqrt{1-\varepsilon^2}\left(\sigma_0 + \sum_{n=1}\sigma_n P_n(\varepsilon)\right), \tag{74}$$

with expansion coefficients σ_n. The expansion coefficients are related to the moments of the density of states Eq.(29) by writing the Chebyshev polynomials explicitly in the form of polynomials with coefficients p_{mk},

$$P_m(\varepsilon) = \sum_{k=0}^{m} p_{mk}\varepsilon^k . \tag{75}$$

Then the expansion coefficients σ_m are obtained in terms of the moments $\mu_{i\alpha}^{(k)}$,

$$\sigma_m = \int_{-1}^{+1}\sum_{k=0}^{m} p_{mk}\varepsilon^k n_{i\alpha}(\varepsilon)\mathrm{d}\varepsilon = \sum_{k=0}^{m} p_{mk}\int_{-1}^{+1}\varepsilon^k n_{i\alpha}(\varepsilon)\mathrm{d}\varepsilon = \sum_{k=0}^{m} p_{mk}\hat{\mu}_{i\alpha}^{(k)}, \tag{76}$$

where we introduced the normalized moments

$$\hat{\mu}_{i\alpha}^{(n)} = \frac{1}{(2b_{i\infty})^n}\sum_{l=0}^{n}\binom{n}{l}(-a_{i\infty})^{(n-l)}\mu_{i\alpha}^{(l)} . \tag{77}$$

Therefore, by calculating the moments $\mu_{i\alpha}^{(k)}$ according to Eq.(29) by pathcounting and inserting the expansion coefficients σ_n into the expansion Eq.(74), one obtains a closed-form approximation of the density of states. Integration of the density of states analytically yields an analytic expression for the bond energy associated with orbital $i\alpha$

$$U_{\mathrm{bond},i\alpha} = \int^{E_\mathrm{F}} (E - E_{i\alpha})\, n_{i\alpha}(\varepsilon)\mathrm{d}\varepsilon = \sum_n \sigma_n \left[\hat{\chi}_{n+2}(\phi_\mathrm{F}) - \gamma\hat{\chi}_{n+1}(\phi_\mathrm{F}) + \hat{\chi}_n(\phi_\mathrm{F})\right], \tag{78}$$

where we introduced the so-called response functions

$$\hat{\chi}_n(\phi_\mathrm{F}) = \frac{1}{\pi}\left(\frac{\sin(n+1)\phi_\mathrm{F}}{n+1} - \frac{\sin(n-1)\phi_\mathrm{F}}{n-1}\right), \tag{79}$$

and the Fermi phase $\phi_\mathrm{F} = \cos^{-1}(E_\mathrm{F}/2b_{i\infty})$.

The lowest order approximation of the analytic bond-order potential that includes only two moments is similar to the Finnis-Sinclair potential[32], so that the analytic BOP expansion may be viewed as a systematic extension of the Finnis-Sinclair potential to include higher moments. On the other hand, as the expression for the bond energy may be integrated analytically for an arbitrary number of moments, the expansion Eq.(78) provides an effective interatomic interaction that may be systematically converged with respect to the exact solution of the TB Hamiltonian by including higher moments. As in the case of the numerical bond-order potentials, the bond energy, Eq.(78), may be rewritten as an equivalent intersite representation. A detailed derivation of the analytic bond-order potentials for transition metals may be found in Ref. 14.

9 Calculation of Forces

The computationally fast and efficient calculation of forces is important for efficient molecular dynamics simulations. In self-consistent electronic structure calculations the Hellmann-Feynman theorem[33,34] makes an efficient calculation of forces possible, as only gradients of the Hamiltonian matrix elements need to be evaluated. The contribution of the bond energy to the forces may be written as

$$\boldsymbol{F}_k = \nabla_k U_{bond} = \sum_{i\alpha \neq j\beta} \Theta_{i\alpha j\beta} \nabla_k H_{j\beta i\alpha} \,. \tag{80}$$

The Hamiltonian matrix elements are pairwise functions, therefore the calculation of the gradients is very efficient. For the bond-order potentials Hellmann-Feynman-like forces[14] may be derived that may be written in a form similar to the Hellmann-Feynman forces Eq.(80),

$$\boldsymbol{F}_k = \nabla_k U_{bond} = \sum_{i\alpha \neq j\beta} \tilde{\Theta}_{i\alpha j\beta} \nabla_k H_{j\beta i\alpha} \,, \tag{81}$$

where $\tilde{\Theta}_{i\alpha j\beta}$ is an approximate representation of the bond order. Just as in the case of the Hellmann-Feynman forces, the calculation of the forces in the bond-order potentials requires only the calculation of the gradient $\nabla_k H_{j\beta i\alpha}$ and not the differentiation of a complex many-body function and is therefore computationally efficient compared to the evaluation of the gradient of an empirical many-body potential.

10 Conclusions

This introductory lecture provides a brief guide to the central ideas and concepts behind the derivation of the bond-order potentials. Instead of diagonalising the TB Hamiltonian, the bond-order potentials provide an approximate local solution of the TB Hamiltonian and the binding energy. The local solution is constructed as a function of the crystal structure or, more general, the positions of the atoms, by relating the electronic structure to the crystal structure using the moments theorem. In this way explicit parametrisations of the energy as a function of the atomic positions are obtained. The accuracy of the bond-order potential with respect to the corresponding tight-binding solution can be improved systematically by including higher moments, which corresponds to taking into account more far-sighted atomic interactions. Hellmann-Feynman-like forces allow for an efficient calculation of the forces in molecular dynamics simulations.

Acknowledgements

We are grateful to D.G. Pettifor for a critical reading of the manuscript. We acknowledge EPSRC funding for part of our work in the projects *Alloy by Design: A materials modelling approach* and *Mechanical Properties of Materials for Fusion Power Plants.*

References

1. A.T. Paxton, *An introduction to the tight binding approximation - implementation by diagonalisation*, in this volume
2. A.P. Sutton, M.W. Finnis, D.G. Pettifor, and Y. Ohta, *The tight-binding bond model*, J. Phys. C **21**, 35, 1988
3. *Modelling electrons and atoms for materials science*, edited by M.W. Finnis and R. Drautz, Prog. Mat. Sci. **52**, Issues 2-3, 2007
4. M.W. Finnis, *Bond-order potentials through the ages*, Prog. Mat. Sci. **52**, 133, 2007
5. M. Aoki, D. Nguyen-Manh, D.G. Pettifor, and V. Vitek, *Atom-based bond-order potentials for modelling mechanical properties of metals*, Prog. Mat. Sci. **52**, 154, 2007
6. R. Drautz, X.W. Zhou, D.A. Murdick, B. Gillespie, H.N.G. Wadley and D.G. Pettifor, *Analytic bond-order potentials for modelling the growth of semiconductor thin films*, Prog. Mat. Sci. **52**, 196, 2007
7. M. Mrovec, M. Moseler, C. Elsässer, and P. Gumbsch, *Atomistic modeling of hydrocarbon systems using analytic bond-order potentials*, Prog. Mat. Sci. **52**, 230, 2007
8. M. Mrovec, D. Nguyen-Manh, D.G. Pettifor, and V. Vitek, *Bond-order potential for molybdenum: Application to dislocation behaviour*, Phys. Rev. B **69**, 094115, 2004
9. D.G. Pettifor, *The tight-binding bond model*, Commun. Phys. (London) **1**, 141, 1976
10. A.R. Mackintosh and O.K. Andersen, Chap. 5.3, *Electrons at the Fermi surface*, (Cambridge University Press, 1980)
11. R. Drautz, D.A. Murdick, D. Nguyen-Manh, X.W. Zhou, H.N.G. Wadley, and D.G. Pettifor, *Analytic bond-order potential for predicting structural trends across the sp-valent elements* Phys. Rev. B **72**, 144105, 2005
12. D.G. Pettifor, *Bonding and Structure of Molecules and Solids* (Oxford University Press, Oxford, 1995)
13. F. Cryot-Lackmann, *On the electronic structure of liquid transition metals*, Adv. Phys. **16**, 393, 1967
14. R. Drautz and D.G. Pettifor, *Valence-dependent analytic bond-order potential for transition metals*, Phys. Rev. B **74**, 174117, 2006
15. P.E.A. Turchi, *Interplay between local environment effect and electronic structure properties in close packed structures*, Mat. Res. Soc. Symp. Proc. **206**, 265, 1991
16. T. Hammerschmidt, B. Seiser, R. Drautz, and D.G. Pettifor, *Modelling topologically close-packed phases in superalloys: Valence-dependent bond-order potentials based on ab-initio calculations*, in: *Superalloys 2008*, edited by R. C. Reed (The Metals, Minerals and Materials Society, Warrendale, 2008), p. 0487
17. F. Ducastelle and F. Cyrot-Lackmann, *Moments developments- II. Application to the crystalline structures and the stacking fault energies of transition metals* J. Phys. Chem. Solids **32**, 285, 1971
18. R. Haydock, *Recursive solution of the Schrödinger equation*, Comp. Phys. Comm. **20**, 11, 1980
19. C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Natl. Bur. Stand. **45**, 225, 1950
20. R. Haydock, V. Heine, and M.J. Kelly, *Electronic structure based on the local atomic environment for tight-binding bands*, J. Phys. C: Sol. Stat. Phys. **5**, 2845, 1972

21. D.G. Pettifor, *New many-body potential for the bond-order*, Phys. Rev. B **63**, 2480, 1989
22. A. Horsfield, A.M. Bratkovsky, M. Fearn, D.G. Pettifor, and M. Aoki, *Bond-order potentials: Theory and implementation*, Phys. Rev. B **53**, 12694, 1996
23. M. W. Finnis, *Interatomic forces in condensed matter* (Oxford University Press, Oxford, 2007)
24. M. Aoki, and D.G. Pettifor, in *International Conference on the Physics of Transition Metals: Darmstadt, Germany, July 2024, 1992*, edited by P. M. Oppeneer and J. K. Kübler (World Scientific, Singapore, 1993), p. 299
25. M. Aoki, *Rapidly convergent bond order expansion for atomistic simulations*, Phys. Rev. Lett. **71**, 3842, 1993
26. D.G. Pettifor, and I.I. Oleinik, *Bounded Analytic Bond-Order Potentials for σ and π Bonds* , Phys. Rev. Lett. **84**, 4124, 2000
27. D.G. Pettifor, and I.I. Oleinik, *Analytic bond-order potentials beyond Tersoff-Brenner. I. Theory* , Phys. Rev. B **59**, 8487, 1999
28. P. Alinaghian, P. Gumbsch, A.J. Skinner, and D.G. Pettifor, *Bond order potentials: a study of s- and sp-valent systems*, J. Phys.: Cond. Mat. **5**, 5795, 1993
29. J. Tersoff, *New empirical model for the structural properties of silicon*, Phys. Rev. Lett. **56**, 632, 1986
30. O.K. Andersen, W. Klose, and H. Nohl, *Electronic structure of Chevrel-phase high-critical-field superconductors*, Phys. Rev. B **17**, 1209, 1978
31. R. Haydock, *Recursive solution of Schrödinger's equation*, Sol. Stat. Phys. **35**, 215, 1980
32. M.W. Finnis, and J.E. Sinclair, *A simple empirical n-body potential for transition metals*, Phil. Mag. A **50**, 45, 1984
33. H. Hellmann, in: *Einführung in die Quantenchemie*, (Deuticke, Leipzig, 1937)
34. R.P. Feynman, *Forces in molecules*, Phys. Rev. **56**, 340, 1939

Coarse Grained Electronic Structure Using Neural Networks

Jörg Behler

Lehrstuhl für Theoretische Chemie
Fakultät für Chemie und Biochemie
Ruhr-Universität Bochum, 44780 Bochum, Germany
E-mail: joerg.behler@theochem.rub.de

The accuracy of the results obtained in theoretical simulations critically depends on the reliability of the employed interatomic potentials. While efficient electronic structure methods like density functional theory (DFT) have found a wide application in molecular dynamics simulations of comparably small systems containing up to a few hundred atoms, for an investigation of many interesting questions one has to deal with systems too large for DFT. In recent years artificial neural networks (NN) have become a promising new tool for the representation of potential-energy surfaces (PES). Due to their flexibility they are able to accurately reproduce a given set of electronic structure data, while the resulting continuous NN-PES can be evaluated several orders of magnitude faster than the underlying electronic structure calculations. Additionally, analytic energy gradients are readily available making NN potentials well suitable for applications to large-scale molecular dynamics simulations. The main drawback of NN potentials is their intrinsically non-physical functional form. Consequently, large reference data sets from electronic structure calculations have to be available to construct reliable NN potentials, which are thus more costly to construct than conventional empirical potentials.

1 Introduction

The investigation of many interesting chemical problems requires long simulations of large systems containing hundreds or thousands of atoms. While in principle accurate electronic structure methods are available[1,2], a direct combination with molecular dynamics (MD) simulations "on-the-fly" is feasible only for small systems, and an application of these methods to large systems is in most cases prohibitively expensive. This dilemma is often circumvented by employing a so-called "divide and conquer" approach. In this approach the costly evaluation of accurate energies and forces by sophisticated electronic structure methods is separated from the actual simulation by a three step procedure. First, the potential is evaluated for a set of representative atomic configurations by highly accurate methods (sometimes also experimental data is used). In the second step a continuous potential representation is constructed, which can be evaluated much faster but should ideally provide essentially the same description as the underlying electronic structure methods. This potential then provides fast access to the potential-energy surface (PES). Finally, in the third step the simulations are carried out employing this potential, which typically allows an extension of length and/or time scales by many orders of magnitude. The use of this approach is wide-spread, a well-familiar example is the use of classical force fields[3–7] in MD simulations, and countless other empirical potentials of varying form and complexity have been developed in the past years for many types of systems.

Fitting complex potential-energy surfaces is a highly non-trivial task. The functional form has to be sufficiently flexible to adapt to the reference points with high accuracy, the

obtained PES should have continuous derivatives for applications in molecular dynamics simulations, it should be fast to evaluate, and its construction should not require a significant amount of manual work. Finally, an improvement and extension in certain regions of the configuration space should be possible without much effort, i.e., without starting the whole elaborate construction of a functional form and the fitting process right from the beginning, if new data points become available. The ideal method also would not be constrained to be applicable only to a certain type of system, like to certain classes of molecules or solids.

For simple systems analytical functional forms containing a few parameters can be "guessed", if possible based on physical knowledge about the system, and the parameters can be fitted to theoretical and/or experimental data. The choice of the functional form requires great care. If the functional form is not chosen appropriately, either unphysical artefacts may be introduced, or the shape of the PES is too much constrained and no accurate fit can be obtained. A well-known example for an analytic fit is the Lennard-Jones 12-6 potential[8]

$$V(r) = 4\epsilon \left\{ \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right\} \tag{1}$$

with two parameters ϵ and σ. This functional form can well represent the interaction between two noble gas atoms, but there is no hope to describe complex organic molecules with this type of pair potential without any angular dependence. More elaborate potentials have been developed for many different types of systems. Classical force fields[3–7] are frequently used for large organic molecules, in particular for studies in the field of biochemistry. Their main drawback, the inability to describe the breaking and making of chemical bonds has been overcome for some systems by so-called "reactive force fields"[9–13]. Many other types of analytic potentials have been developed in recent years, and discussing even a small fraction is beyond the scope of the present lecture.

The general advantage of analytical fits is that the number of parameters is rather small, consequently only a few reference calculations are sufficient for setting up the potentials. Additionally, the individual terms often allow for a physical interpretation and an unexpected behavior of the potentials during simulations is rare. Problems may arise, if the functional form is too simple and thus cannot reproduce the reference data. For many systems the construction of suitable functional forms has been a frustrating challenge, and the sheer complexity of chemical bonding enforces either a limitation to a subset of possible structures that can be accurately described (e.g. by prohibiting bond breaking) or the chemical complexity has to be reduced. Most "reactive" potentials developed so far are thus applicable only to monocomponent or binary systems and have been mainly applied in the field of materials science[14–17] or to describe low dimensional PESs in surface science[18–20].

Another possibility to construct PESs is to use purely mathematical fitting methods like splines. They have a very general functional form, but they are not applicable to high-dimensional PESs, because the error increases rapidly with the number of dimensions, and they are very sensitive to noise in the data. An example for a very general approach for the description of molecule-surface interactions is the modified Shephard method[21,22], which is based on a Taylor expansion of the energy around the reference points.

In recent years, another mathematical approach based on artificial neural networks

(NN) has become a promising new tool for the construction of potential-energy surfaces. Artificial neural networks have been first introduced in 1943 by McCulloch and Pitts inspired by the way of signal processing in the brain[23]. The robustness, fault tolerance, ability to handle noisy data and highly parallel structure of the brain since then stimulated a lot of research and many attempts have been made to mimic these properties in computer algorithms. Nowadays, neural networks have found numerous applications in many fields of science. They have the general ability to associate some input pattern with a response, and are very well suited for applications like finger print identification, voice recognition and many others[24]. Apart from pattern recognition applications they can also be used to approximate functions based on a known set of function values[25].

There are various examples for applications in chemistry[26–29]: establishing structure-activity relationships[30], prediction of reaction probabilities[31], medical classification problems in clinical chemistry[32], binding site prediction[33], extraction of pair potentials from diffraction data[34], estimation of force constants in large molecules[35], "data mining" in drug design[36], electrostatic potential analysis[30,29], construction of exchange correlation potentials[37], the numerical solution of the Schrödinger equation for simple model systems[38], the prediction of secondary protein structure[39], and the detection of signals in the presence of noise[40].

The topic of this lecture is the application of neural networks to the construction of potential-energy surfaces employing their ability to approximate unknown functions very accurately. The central assumption behind the construction of potential-energy surfaces using neural networks is that an analytic representation of the PES exists. This analytic form may be very complex (indeed too complex to be ever written down) and completely unknown. But if it in principle exists, the neural network can be used to approximate this unknown functional form to high accuracy, since it has been proven that any real-valued function depending on a set of variables can be represented by a feed-forward neural network with arbitrary precision[25,41,42]. The reason for this capability of NNs is the extreme flexibility arising from a large number of simple non-linear functions which are combined in a hierarchical and systematic way. This is similar to the interconnection of biological neurons in the nervous system and gave the method its name. The structure of NNs is described in detail in the following sections.

Translating this capability of NNs to real world applications is not straightforward and a lot of effort has been put into the construction of neural network potentials by many research groups in chemistry and physics. The aim of this lecture is not to give a complete review of all techniques developed so far, but to point out some conceptual problems and how they can be solved for different types of chemical applications. For this purpose we will in particular focus on the class of multilayer feed-forward neural networks, which is most important for the representation of potential-energy surfaces.

2 Neural Network Potentials

2.1 The Functional Form of a Feed-Forward Neural Network

Since the introduction of artificial neural networks many different NN types have been developed[24,43–45]. For the representation of potential-energy surfaces in particular feed-forward neural networks have gained a central role. They have a very general, i.e. unbiased, form, which is a nested function of a large number of simple functional units. The

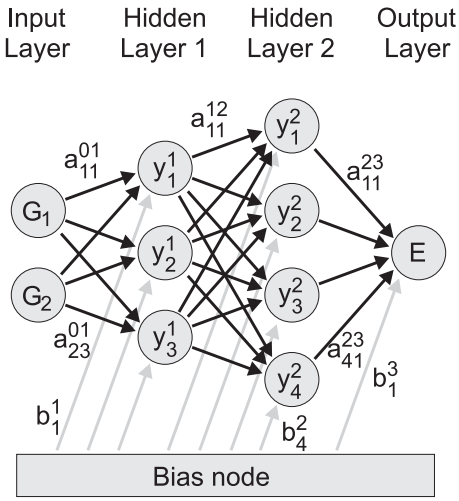


Figure 1. Schematic structure of a simple 2-3-4-1 feed-forward neural network. The value of the node in the output layer corresponds to the energy E, which depends on the variables G_i in the input nodes. The bias weights b_i^j connect the bias node with the nodes i in layer j and shift the non-linear region of the activation functions. The number of hidden layers and the nodes in these layers define the functional form of the NN. The weights parameters a_{ij}^{kl} connecting node i in layer k with node j in layer l as well as the bias weights are the fitting parameters of the NN. They are determined iteratively using a set of known reference data points.

functional form of such a NN can be visualized schematically as shown for a small example NN in Fig. 1. The NN consists of artificial neurons called nodes and are represented by the grey circles. These nodes correspond to the neurons in the biological NN. They are arranged in several layers: an input layer, an output layer, and one or more hidden layers. The nodes in the input layer correspond to the i variables G_i of the potential-energy function, i.e., to the atomic degrees of freedom that determine the total energy of the system. The node in the output layer is the target quantity, the potential-energy of the system. The purpose of the remaining parts of the NN is to set up a functional relation between the atomic positions and the energy. The specific functional form is defined by the number of hidden layers and the number of nodes in each layer. The term "hidden layer" indicates that the nodes in these layers have no physical meaning and are just auxiliary mathematical constructs to provide the required flexibility. Each node in each layer is connected to the nodes in the adjacent layers by so-called weight parameters, the fitting parameters of the NN. Here, a_{ij}^{kl} is the weight parameter connecting node i in layer k with node j in layer l. The weight parameters are represented by the arrows in Fig. 1. Additionally, the nodes in the hidden layers and the output node are connected to a bias node via bias weight parameters. We use the symbol b_i^j, which is the bias weight acting on node i in layer j.

Once the topology of the NN is set up, the output is calculated in the following way: First, the numerical values of the coordinates of the atoms in a given structure are provided to the NN in the nodes of the input layer, which has the layer index "0". The value G_i of each node i is then propagated to each node in the first hidden layer and multiplied by the value of the connecting weight parameter. At each node j in the first hidden layer the sum of products $\sum_i G_i a_{ij}^{01}$ is calculated. So far this corresponds to a linear combination of

the atomic coordinates. Clearly, in the true PES there is a complicated non-linear relation between the energy and the atomic positions, and the capability to represent these general non-linear functions is introduced by applying a non-linear activation function f_j^k to the final sum at each node j in layer k. Examples for activation functions will be given below. Frequently used activation functions have a sigmoidal shape, i.e., they have a finite non-linear region and saturate for very small and very large arguments. The role of the bias weights b_i^j is to shift the sum at each node into the non-linear regime of the activation functions (cf. Sec. 2.2). In summary, the value y_j^1 of node j in the first hidden layer is then calculated by

$$y_j^1 = f_j^1 \left(\sum_i b_j^1 + a_{ij}^{01} G_i \right) \tag{2}$$

and for a general node j in layer k the equation becomes

$$y_j^k = f_j^k \left(\sum_i b_j^k + a_{ij}^{k-1,k} y_i^{k-1} \right) \quad . \tag{3}$$

The number obtained at each node in the first hidden layer is then propagated to each node in the second hidden layer and again multiplied by the respective connection weight. At each node in the target layer again an activation function is applied and so forth until finally a number is obtained at the node in the output layer of the NN.

The full functional form of the example NN in Fig. 1 is given accordingly by

$$E = y_1^3 = f_1^3 \left(b_1^3 + \sum_{k=1}^{4} a_{k1}^{23} \cdot f_k^2 \left(b_k^2 + \sum_{j=1}^{3} a_{jk}^{12} \cdot f_j^1 \left(b_j^1 + \sum_{i=1}^{2} a_{ij}^{01} \cdot G_i \right) \right) \right) \tag{4}$$

In general the NN output depends on the topology of the NN, i.e., the number of layers and nodes per layer, the type of activation functions, and most importantly, the numerical values of the weight parameters. Initially, the weight parameters are chosen randomly and the output of the NN is of course very different from the correct potential-energy of the structure. But if for a set of reference structures the potential-energy is known, e.g. from electronic structure calculations, then an error function can be defined as the difference between the output of the NN and the known correct energy. This error function can then be minimized by optimizing the weight parameters until all example points are accurately reproduced by the NN. Details on the weight optimization are given in Section 3 below. This optimum set of weight parameters is then kept fixed and can be used to predict the energies of new (similar) structures not included in the reference set, for instance structures visited in the trajectory of a MD simulation.

The size of the NN is determined empirically for a given system by constructing fits with different numbers of hidden layers and nodes per layer, and choosing the structure which provides the most accurate fit. Care has to be taken in that large networks may contain too many parameters. The resulting high flexibility may yield overfitting, which has to be checked carefully as will be discussed in Sec. 3.4. As a general rule, if two NN architectures provide the same accuracy the one with less parameters should be preferred. It has also been suggested to adapt the number of nodes during the fit, e.g. by employing genetic algorithms[46], but due to the increased computational costs of this approach so far it did not find regular use in the context of potential-energy surfaces.

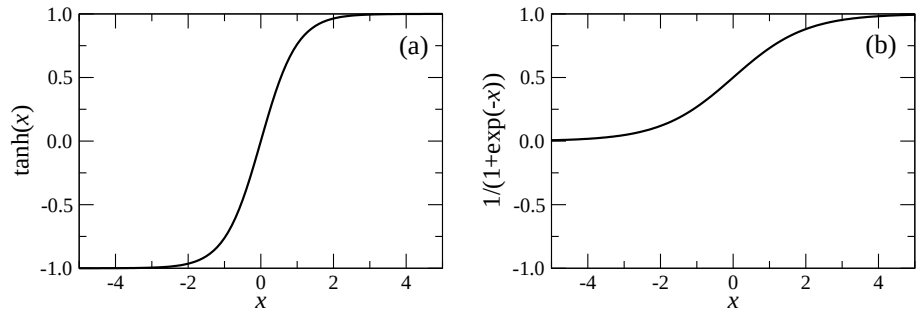


Figure 2. Frequently used activation functions: hyperbolic tangent (a), and sigmoid function (b). Activation functions saturate for very small and very large arguments, but have a non-linear region in between, which ensures that the neural network is able to adapt to general non-linear functions.

The topology of a feed-forward neural network can be described by a set of numbers defining the number of nodes in the input layer, each hidden layer and the output layer[47]. The simple example network in Fig. 1 in this scheme is a 2-3-4-1 NN. Additional letters can be provided describing the type of activation functions used in the individual layers, e.g. t for a hyperbolic tangent, s for a sigmoid function, and l for a linear function (cf. Section 2.2).

2.2 Activation Functions

Neural networks obtain the ability to fit general, i.e., non-linear, functions by the incorporation of so-called activation functions. Activation functions are also called "transfer functions" or "basis functions" of the network. In general they map a variable x to a range between -1 and 1 or between 0 and 1. This is a consequence of their general property that they saturate to these numbers for very small and very large values of x and have a non-linear region in between. Frequently used examples for activation functions are the sigmoid function

$$f(x) = \frac{1}{1+e^{-x}} \quad , \tag{5}$$

the Gaussian function

$$f(x) = e^{-\alpha x^2} \quad , \tag{6}$$

or the hyperbolic tangent

$$f(x) = tanh(x) \quad . \tag{7}$$

The sigmoid function and the hyperbolic tangent activation functions are plotted in Fig. 2. For the output node sometimes also a linear function is used to avoid any constraint on the possible range of output values,

$$f(x) = x \quad . \tag{8}$$

Alternatively, if e.g. a hyperbolic tangent is applied, the range of energy values can be rescaled before the fitting to the interval between -1 and 1 that corresponds to the range of

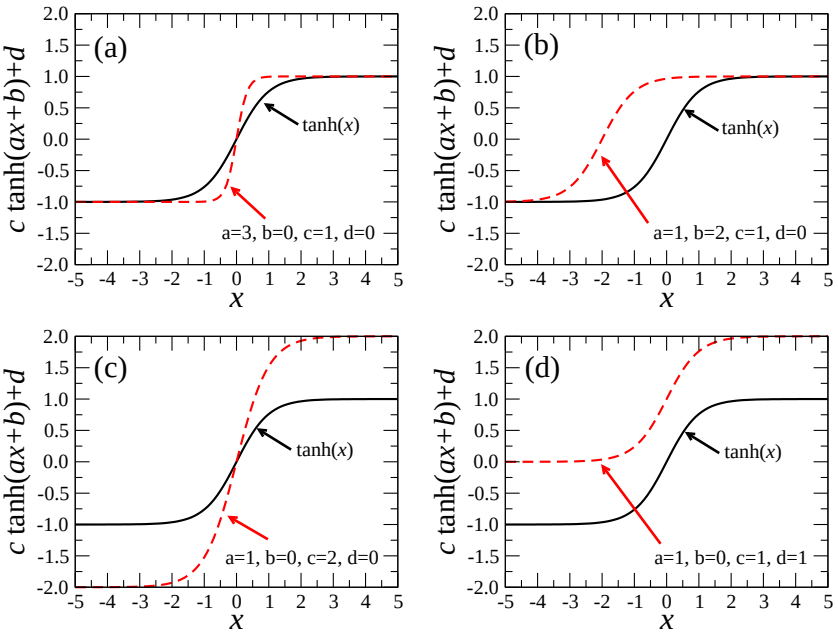


Figure 3. Illustration of the flexibility of the hyperbolic tangent activation function. The functional form of a neural network (Eq. 4) contains building blocks $f(x) = d + c \cdot \tanh(a \cdot x + b)$ which can adapt to general functions by varying the parameters a, b, c, and d. For comparison also the unmodified hyperbolic tangent is plotted as black line.

values of the hyperbolic tangent, and the values at the output node are scaled back to the original range. The activation functions have to be continuous and differentiable, which is needed for the application of standard optimization algorithms, but also for the calculation of the derivatives of the output with respect to the atomic coordinates, i.e., for the atomic forces.

We will illustrate the capability of the non-linear activation functions to adapt to arbitrary functions using the example of the potential of the harmonic oscillator

$$V(x) = x^2 \quad . \tag{9}$$

For instance for the hyperbolic tangent activation function the nested form of the neural network energy expression (e.g. Eq. 4) can be decomposed into a set of functional units of the form

$$f(x) = d + c \cdot \tanh(a \cdot x + b) \tag{10}$$

with four "parameters" a, b, c, and d. By optimizing these parameters, the shape of the hyperbolic tangent can be modified as illustrated in Fig. 3 This flexibility can be used to obtain a rather good approximation to the parabolic potential in a given range by just 2 activation functions, as shown in Fig. 4. Finally, we note that it has also been suggested to employ periodic activation functions, which can facilitate fitting periodic functions like torsional potentials[48].

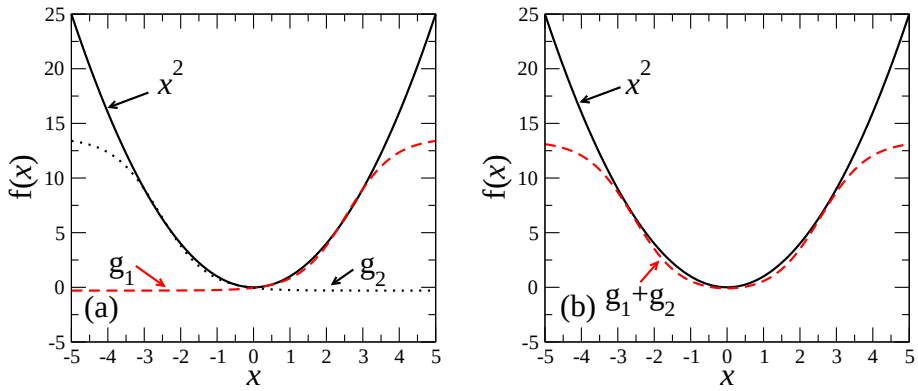


Figure 4. Example for the fit of a parabola in the range $-3 < x < 3$ by two hyperbolic tangent functions g_1 and g_2. In (a) the two functions are plotted separately, in (b) the sum is compared with the parabola.

2.3 Symmetry Functions

For several reasons it is advantageous not to use directly the Cartesian coordinates of the atoms as input for the neural network but to perform a transformation to coordinates which are physically more appropriate. These new sets of coordinates are often called "symmetry functions"[49]. This transformation is necessary because the numerical values of the Cartesian coordinates of the atoms do not carry the structural information defining the energy of the system in a directly usable form, but the relevant information is included in the relative positions of the atoms with respect to each other. A mere translation or rotation of the complete system must not change its energy, but it clearly affects the numerical values of the Cartesian coordinates. If these coordinates were used as input for the NN, the NN output would change with translation or rotation. A very basic coordinate transformation to avoid this problem is to switch from Cartesian to internal coordinates, i.e., defining the system in terms of bond lengths, bond angles and torsion angles. We note here that in contrast to classical force fields neural networks in general do not require the the specification of bonds ("bonded" atom pairs) and angles, so these terms could also be called two-, three- and four-body terms.

The use of internal coordinates provides a reasonable description of small molecular systems without any significant structural change like the breaking of bonds. Larger molecules, however, will contain many atoms of the same chemical species, and if two atoms of the same element are simply exchanged, the total energy of the system must not change. This symmetry information is not included in a set of internal coordinates, because the order in which the internal coordinates are fed into the NN is not arbitrary. Capturing this additional symmetry for general systems is a difficult task. For low-dimensional systems a sequence of symmetrization and antisymmetrization steps has been suggested[50,51]. To illustrate the procedure, imagine a water molecule with the bond lengths $r_{\mathrm{OH},1}$ and $r_{\mathrm{OH},2}$. A proper set of input coordinates for the NN has to take into account that both hydrogen atoms are indistinguishable, i.e., their numbering is arbitrary and exchanging the interatomic distances $r_{\mathrm{OH},1}$ and $r_{\mathrm{OH},2}$ must not change the output of the NN. By squaring the sum and the difference of $r_{\mathrm{OH},1}$ and $r_{\mathrm{OH},2}$ we obtain two functions G_1 and G_2, which

are independent of the numbering of the hydrogen atoms, i.e., exchanging both atoms does not change the values of G_1 and G_2.

$$G_1 = (r_{\mathrm{OH},1} + r_{\mathrm{OH},2})^2 \tag{11}$$

$$G_2 = (r_{\mathrm{OH},1} - r_{\mathrm{OH},2})^2 \tag{12}$$

In order to completely define this system with three degrees of freedom, a third coordinate for the distance between both hydrogen atoms or the angle H_1-O-H_2 can be introduced, which does not need to be symmetrized.

For an extension of this symmetrization/antisymmetrization scheme to high-dimensional systems soon numerical problems arise from the emergence of very large and very small numbers for the G_i as well as from the increasingly complicated terms. This can be seen from the example methane, for which we give here the symmetrized terms for the four CH distances.

$$G_1 = \left[(r_{\mathrm{CH},1} + r_{\mathrm{CH},2})^2 + (r_{\mathrm{CH},3} + r_{\mathrm{CH},4})^2\right]^2 \tag{13}$$

$$G_2 = \left[(r_{\mathrm{CH},1} + r_{\mathrm{CH},2})^2 - (r_{\mathrm{CH},3} + r_{\mathrm{CH},4})^2\right]^2 \tag{14}$$

$$G_3 = \left[(r_{\mathrm{CH},1} - r_{\mathrm{CH},2})^2 + (r_{\mathrm{CH},3} - r_{\mathrm{CH},4})^2\right]^2 \tag{15}$$

$$G_4 = \left[(r_{\mathrm{CH},1} - r_{\mathrm{CH},2})^2 - (r_{\mathrm{CH},3} - r_{\mathrm{CH},4})^2\right]^2 \tag{16}$$

Many more symmetrized terms for the H-H distances and/or HCH angles are required to describe all degrees of freedom. It is immediately obvious that this scheme cannot be extended to larger molecules containing many atoms of the same species. Nevertheless, for low-dimensional potential-energy surfaces this scheme is very useful and can also be applied to molecule-surface interactions. In the latter case further compilations are related to the lateral periodicity of crystalline surfaces. For the construction of symmetry functions in this case the bond lengths have to be replaces by Fourier terms reflecting the surface symmetry[51].

In general, the choice of symmetry functions is system-dependent, but they have to fulfill several requirements. They need to be continuous in value and slope, they should be fast to evaluate, and there should be a one-to-one correspondence between a given structure and its set of symmetry function values. If two structures with different energies yield the same set of symmetry function values, fitting the NN is not possible, because the NN would associate two different energies to the same structure. It should also be noted that there is no need to ever invert the transformation of the coordinates. The mapping is always from atomic configurations to symmetry functions, for the construction of the training set as well as for the energy prediction of a new structure.

Typically the range of values of each symmetry function is scaled. This has numerical reasons, because it is advantageous to avoid symmetry function values in the saturation region of the activation functions. In this case the function values of the activation functions would be about the same for all symmetry function values. Further, depending on the definition, symmetry functions may have a very large or very small range of values, in particular if symmetrization/antisymmetrization is applied. Also in this case it is advantageous to rescale the range of values to an interval between 0 and 1.

In summary, it is usually not possible to use the Cartesian coordinates of the atoms to construct a NN PES. Instead, in a first step, the Cartesian coordinates are mapped onto a

set of symmetry functions $\{G_i\}$. In the next step the symmetry function values are used as input for the NN, which then yields the energy of a structure.

2.4 Atomic Forces

The analytic form of the neural network total energy expression in Eq. 4 allows to calculate analytic derivatives, which are essential to obtain accurate forces for applications in MD simulations. If an intermediate mapping of the atomic coordinates onto symmetry functions $\{G_i\}$ is used, the force with respect to an atomic coordinate α is given by

$$\begin{aligned} F_\alpha &= -\frac{\partial E}{\partial \alpha} \\ &= -\sum_i \frac{\partial E}{\partial G_i} \frac{\partial G_i}{\partial \alpha} \end{aligned} \tag{17}$$

The derivative $\frac{\partial E}{\partial G_i}$ is given by the NN topology, the derivative $\frac{\partial G_i}{\partial \alpha}$ is defined by the choice of symmetry functions. Also other quantities containing analytic derivatives like the stress tensor are directly accessible.

3 Optimization of the Weight Parameters

3.1 The Fitting Procedure

In order to predict the potential-energy of an atomic configuration, the weight parameters of the NN have to be known. Typically, these parameters are optimized iteratively using a set of known function values. This optimization process is called "training" or "learning" of the NN.

A large variety of algorithms can be used to optimize the weight parameters[52], which can be classified as gradient-based algorithms and random methods. Examples for gradient-based methods are the steepest-descent algorithm, which is called "back-propagation" in the NN community, conjugate gradients[53–55], the global extended Kalman filter[56], and many more. Gradient-based learning schemes are likely to get trapped in local minima at some point, but they are fast. Examples for random methods are the weight optimization employing genetic algorithms[52] or a swarm search[52]. Random methods can easily jump from one local minimum to another, but they are computationally very demanding. A method combining ideas from gradient-based and random methods is simulated annealing[57,35,52], which is essentially a damped dynamics in the space of the weight parameters.

For complex potential-energy surfaces and large data sets, which can easily reach the order of 10000 reference points, typical NNs contain between one and three hidden layers and between 25 and 40 nodes per layer. Consequently, roughly 1000 to 5000 weight parameters are used. The optimization of such a large number of weight parameters is a formidable task and there is no hope in practical fits to find the global minimum. Still, many local minima may represent the training set sufficiently well and can provide a reliable NN potential, and often many fits of a comparable quality are found with different sets of weight parameters. The resulting NNs, which yield about the same fit quality but cannot be transferred into each other by a simple permutation of the NN nodes, are called degenerate NNs[58].

The optimization of the weight parameters corresponds to the minimization of a cost function Γ, which is defined as the sum of the squared errors of the energies $E_{i,\mathrm{NN}}$ predicted by the NN and the "true" reference energies $E_{i,\mathrm{Ref}}$ from electronic structure calculations.

$$\Gamma = \sum_{i=1}^{N} \frac{1}{N} (E_{i,\mathrm{NN}} - E_{i,\mathrm{Ref}})^2 \tag{18}$$

It is also possible to modify this cost function by assigning different fitting weights (not to be confused with the weight parameters) to enforce a more accurate fit for certain parts of the PES, e.g. along the reaction path[49].

The optimization process is started by initializing the weight parameters as random numbers, typically in the range $[-1, 1]$. In each iteration, which is also called "epoch" in the NN community, each training point is presented once to the NN. Usually the training points are presented in random order, to reduce the probability of ending up in a close local minimum.

The measure for the quality of a fit is the root mean squared error of the training points

$$RMSE = \sqrt{\frac{1}{N} \sum_{i} (E_{i,\mathrm{NN}} - E_{i,\mathrm{Ref}})^2} \tag{19}$$

which is calculated in every epoch. Sometimes also the mean absolute deviation

$$MAD = \frac{1}{N} \sum_{i} |E_{i,\mathrm{NN}} - E_{i,\mathrm{Ref}}| \tag{20}$$

is monitored. The course of the RMSE in a typical fit will be discussed in Section 3.4.

Because of their efficiency compared to random fitting methods, gradient-based algorithms play a dominant role. For their application the partial derivatives of the NN output

$$\frac{\partial \Gamma}{\partial a_{ij}^{kl}} = \sum_{\mu} \frac{1}{\mu} \frac{\partial E_{\mu,\mathrm{NN}}}{\partial a_{ij}^{kl}} \tag{21}$$

and

$$\frac{\partial \Gamma}{\partial b_{i}^{j}} = \sum_{\mu} \frac{1}{\mu} \frac{\partial E_{\mu,\mathrm{NN}}}{\partial b_{i}^{j}} \tag{22}$$

with respect to all weight parameters have to be calculated for each training point. There are two types of learning. In the so-called "offline" learning the weights are updated once per epoch, e.g. if a conjugate gradient is applied. In "online" learning the weights are updated after the presentation of each individual training point. In this case the summation in Eqns. 21 and 22 are dropped and the gradients for each single point are used separately. An example for an algorithm for online learning is the global extended Kalman filter. In the following Sections two frequently used optimization algorithms, back-propagation and the global extended Kalman filter will be discussed.

3.2 Back-Propagation

The most frequently employed algorithm for the optimization of the weight parameters is "back-propagation". Essentially, this is identical to a standard steepest-descent optimization. In the back-propagation optimization the weight parameters are updated according to

$$a_{ij,new}^{kl} = a_{ij,old}^{kl} - \kappa \frac{\partial \Gamma}{\partial a_{ij}^{kl}} \tag{23}$$

κ is a positive damping factor called "learning rate". The term "back-propagation" has its origin in the way the derivatives with respect to the weights are calculated. The output of a feed-forward NN is calculated in a so-called "forward-pass" through the NN. First, using the values at the input nodes and of the connecting weights the values at the nodes in the first hidden layer are determined. They are then passed forward to the second hidden layer, to evaluate the numerical values at the nodes in this layer and so on. Consequently, the total output of the NN is calculated by propagating the information forward through the NN. On the other hand, for the calculation of the derivatives of the NN output value with respect to the connecting weights the information flow is in the opposite direction, "backwards".

As a steepest-descent method back-propagation is not very efficient and likely to get trapped in a local minimum. It may also show an oscillating behavior or diverge if κ is chosen too large, and the optimum value of κ is system-dependent.

3.3 The Kalman Filter

An optimization scheme which has become very popular in the context of neural networks is the extended Kalman filter. The global extended Kalman filter is a very sophisticated algorithm originating from estimation and control theory[59]. It is used for online learning, i.e., the weight parameters are optimized after the presentation of each individual training point. In the Kalman filter the update is directed by a weighted history of previous updates of the weight parameters. The derivation of the equations used in the weight update is beyond the scope of the present lecture, and here we will just present the result for the Kalman filter recursion relations for the update n:

$$\mathbf{K}(n) = \lambda^{-1}\mathbf{P}(n-1)\mathbf{J}(n)\left[\mathbf{I} + \lambda^{-1}\mathbf{J}^T(n)\mathbf{P}(n-1)\mathbf{J}(n)\right]^{-1} \tag{24}$$

$$\mathbf{P}(n) = \lambda^{-1}\mathbf{P}(n-1) - \lambda^{-1}\mathbf{K}(n)\mathbf{J}^T(n)\mathbf{P}(n-1) \tag{25}$$

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{K}(n)\left[\mathbf{E}_{\mathrm{Ref}}(n) - \mathbf{E}_{\mathrm{NN}}(\mathbf{w}(n-1)\right] \tag{26}$$

$\mathbf{K}$ is the Kalman gain matrix, and $\mathbf{J}$ is the Jacobi matrix with the elements

$$J_i = \frac{\partial E}{\partial w_i} \tag{27}$$

where w is either a connection or a bias weight parameter. $\mathbf{P}$ is the covariance matrix, and $\mathbf{I}$ is the identity matrix. For each training point first the Kalman gain matrix is updated using the covariance matrix of the previous epoch and the current weight derivatives in the Jacobi matrix. Then the new vector of weight parameters $\mathbf{w}$ is determined using $\mathbf{K}$. Finally, the covariance matrix is updated according to Eq. 25. A "forgetting schedule" is

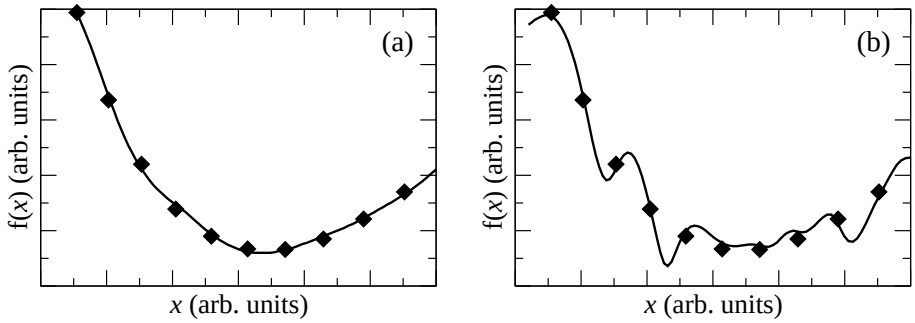


Figure 5. Illustration of "overfitting". In (a) the training points (diamonds) are well represented and smoothly connected by the fit (line). Also in (b) the fit is very accurate for the training points, but many local extrema are present in between the training points. These extrema are not based on information in the training set but are artefacts of the fit arising from too much flexibility. This "overfitting" can be detected for example by using a test set of point located in between the training points. If the error of the test points is significantly higher than the error of the training points, overfitting is present.

introduced via λ to ensure that only the recent history of updates in taken into account for the update of point n,

$$\lambda(n) = \lambda_0 \lambda(n-1) + 1 - \lambda_0 \quad . \tag{28}$$

The constant λ_0 is usually chosen between 0.99000 and 0.99900.

Adapting the weight parameters after each training point is computationally rather costly. In its adaptive form the extended Kalman filter thus is not used to update the weight parameters after each individual training point, but an error threshold α is defined in terms of the actual RMSE of the full training set. Only if the error of a training point is larger than the product of α and the current RMSE, the point will be used to update the weights. This can reduce the computational effort significantly, since only points are used in the fit, which are not well represented.

For the construction of NN potentials, the extended Kalman filter often shows a performance which is superior to other optimization algorithms[56,58], because is less likely to get trapped in shallow local minima.

3.4 Overfitting

Employing a very flexible functional form immediately rises the question, how overfitting can be detected and controlled. If a set of training points is fitted very accurately while other points not included in the training set are poorly described, this is called "overfitting". In other words, overfitting is an improvement of the fit in one region of the configuration space at the cost of a poor quality in another region. This is illustrated in Fig. 5. In (a) the training points are well represented and connected by a smooth curve which seems to be a reasonable fit. The RMSE of the training points will be very low in this case. In (b), however, the RMSE will be very low as well, possibly lower than in (a), because also here the curve is very close to all the training points indicated as diamonds. Nevertheless, in (b) the curve shows many local extrema, which are apparently not justified by the training data. This is a typical example for overfitting.

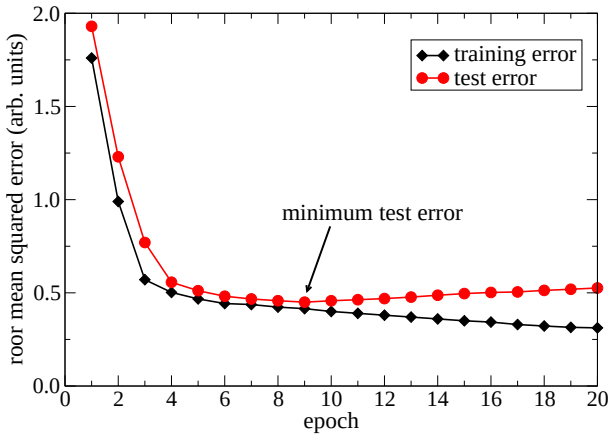


Figure 6. Typical course of the errors of the training set and the test set in the iterative optimization of the NN weight parameters.

Most applications of neural networks do not allow for a visual inspection of the fit quality due to the high dimensionality. A commonly employed method in these cases is the so-called "early stopping" method. In this method the available points are split into a training set which is used to update the weight parameters, and a test set, which is not used in the optimization procedure. A comparison of the RMSEs of the training and the test set then allows for an estimate of the generalization properties of the NN fit. The typical course of the training error and the test error in the iterative fit is shown in Fig. 6. The training error decreases steadily since the weight parameters are optimized to reproduce the training set as accurately as possible. Initially, also the test error drops quickly, because the description of the overall shape of the potential-energy surface is improving in each epoch. Then the test error reaches an local minimum and starts to increase slowly. This increase indicates that now the accuracy of the training points is improved on the expense of the regions in configuration space in between the training points. This is detected by the RMSE of the test points, which are located in between the training points. In the "early stopping" method, the set of weights, which minimizes the error of the test set is considered to represent the best fit.

4 Construction of the Training Set

The training set for the optimization of the weight parameters can be obtained from any electronic structure method, because the only information required for each atomic configuration in the training set is its total energy. Information on energy gradients can in principle also be used in the construction of the NN potentials[58], but in practice this is rarely done. A significant constraint on the choice of the electronic structure method for large systems is the large number of training points that is needed to set up a neural network potential. This limits the application of computationally demanding but very accurate quantum chemical methods to small molecules, and the most frequently used electronic structure method for large systems is density functional theory.

Having chosen an electronic structure method for the calculation of the training set, the next problem is the choice of the atomic configurations. For small systems with few degrees of freedom, e.g. small molecules, a dense grid of points can be obtained by systematically varying all degrees of freedom. However, the exponential growth of the number of configurations prevents a systematic mapping for larger systems of typically more than six degrees of freedom. This is because if each degree of freedom is sampled by N points, for d degrees of freedom the number reference calculations is N^d. In practical applications like MD simulations often only a subspace of the full configuration space is accessible for the system. This relevant subspace can be mapped by a systematic approach in the following way: First, some random structures are calculated and a preliminary NN potential is constructed using these points. This potential will not be reliable and may contain unphysical stationary points. Nevertheless it can be used to perform short MD simulations or structural relaxations to propagate the system to new configurations. The configurations suggested in this way by the NN can then be recalculated using electronic structure methods, and the resulting energies can be compared with the NN predictions. If the agreement is not satisfactory, the new structures can be included in the training set and the fit can be refined. In contrast to conventional empirical potentials with a given analytic form and a few adjustable parameters no change in the functional form of the NN is required, and improving the NN is straightforward without any manual work. The new fits again can be used to suggest new structures and so forth, until all wrong features of the NN PES have been removed and the training data set is reproduced with the desired accuracy. Typically, an accuracy of a few meV per atom with respect to the reference energies can be obtained in this way.

It is also possible, to identify regions of the configuration space, which are relevant but not well represented in the training set, without carrying out costly electronic structure calculations. For this purpose several NN fits are constructed employing different NN topologies. Because the NN topologies are different, so is the functional form of the fits. Now two fits with approximately the same RMSEs for the training and the test set are chosen. Accordingly, it is not possible to judge which of the two fits is a better representation of the true PES. Then a large number of structures is generated, e.g., random structures, optimized structures or snapshots from MD simulations. The energy for all structures is predicted by both fits. If a structure is close to a point already included in the training set, both NNs are likely to predict a similar energy, otherwise the RMSEs of the NNs would be clearly different. But if the predicted energies are very different, then the NNs have too much flexibility at this point in the configuration space and an electronic structure calculation should be carried out for this point. This way it is possible to avoid a large number of unnecessary electronic structure calculations.

5 Applications of Neural Network Potential-Energy Surfaces

5.1 Neural Network Potentials for Molecular Systems

To date, the most frequent application of NN potentials is the representation of rather low-dimensional molecular PESs. Many different sets of coordinates have been developed to transform the atomic positions to suitable inputs for the NN. For simple diatomic molecules the potential depends only on the interatomic distance and one input coordinate is sufficient

for the PES construction. An example is a study of the photodissociation of the HCl^+ ion using NN PESs for the ground and excited state[60]. Also transition probabilities in the OH radical have been studied[61]. However, for larger molecules the coordinates can become significantly more complicated. Ideally, the coordinate transformations takes the molecular symmetry directly into account. The exact form of the applied set of symmetry functions is highly system-specific. In a study of the vibrational levels of the H_3^+ ion a symmetrical formulation has been suggested taking into account the equivalence of all three nuclei[62,63]. NNs have also been used to describe the interaction between two HF molecules as well as between a HF and a HCl molecule[64].

NN potentials can be very useful in situations when the polarizability of molecules complicates the construction of classical potentials. This has been shown for the example of the Al^{3+} ion in water by combining conventional two-body terms with three-body interactions, which are represented by a flexible NN[50]. The equivalence of the two water molecules interacting with the Al^{3+} has been included by symmetrized coordinates. NN potentials have also been constructed for the water dimer[65], and later applied to liquid water in combination with empirical parameters of the TIP4P water model[66]. A very general method for molecules has been suggested based on a high-dimensional model representation employing a many-body expansion of the potential[67–70]. This approach is very systematic and promises a very high accuracy, but due to its complexity and computational demand, it is still limited to a rather small number of degrees of freedom in practical applications.

5.2 Neural Network Potentials for Molecule-Surface Interactions

Neural network potentials have also been applied to the description of molecule-surface interactions[47,49,71]. In contrast to molecular systems the number of degrees of freedom is typically much larger, because realistic surface models for instance in form of periodically repeated slabs contain a significant number of atoms. Two approaches have been followed in order to reduce the resulting complexity: either only a few degrees of freedom of selected surface atoms are included[47], or a frozen surface approximation is applied, i.e., all degrees of freedom of the surface atoms are eliminated by freezing their positions. For diatomic molecules this approach reduces the problem to a six-dimensional potential-energy surface, which can then be mapped systematically on a grid of a few thousand points by electronic structure calculations. The frozen surface approximation is a drastic approximation and its validity has to be checked carefully for each individual system. It has been found that physical quantities, which are less sensitive to a motion of the surface atoms, can be calculated to good accuracy[51,72–74,49].

Applications to molecule-surface interactions require a special type of symmetry functions[71,51], which have to include the periodicity of the surface as well as all symmetry elements of the surface unit cell. A transformation of the molecular coordinates to these symmetry functions is then equivalent to folding the configuration space into the symmetry unique wedge of the surface. An example of the symmetry unique wedge of the (111) surface of an fcc metal is shown in Fig. 7. This significantly reduces the computational costs for the calculation of the reference electronic structure calculations, because only configurations inside the symmetry unique wedge have to be calculated. Whenever the molecule leaves this wedge in the course of the trajectory, its coordinates are folded back

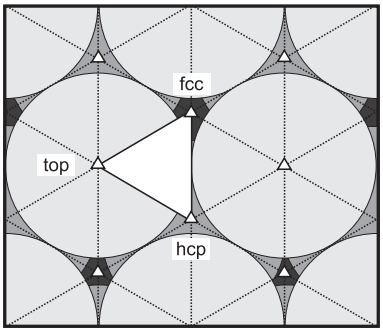


Figure 7. Symmetry unique wedge of the (111) surface of an fcc metal spanned by the top, fcc and hcp sites. The mirror planes perpendicular to the surface are indicated as dotted lines, the triangles represent threefold rotational axes.

to an energetically equivalent position in the symmetry unique wedge. In MD applications the NN PES needs to be continuous in value and slope. A special requirement for the description of molecule-surface interactions is that the symmetry functions have continuous derivatives at the boundaries of the symmetry unique wedge to avoid discontinuities of the atomic forces. A detailed discussion of a scheme for the systematic construction of suitable symmetry functions can be found elsewhere[51].

5.3 High-Dimensional NN Potentials for Condensed Systems

To date NN potentials have been mainly applied to PESs of gas phase molecules with a rather low number of up to about 12 degrees of freedom. For general chemistry in condensed systems, e.g. in solution, in the solid state or at surfaces, an extension of the NN methodology to high-dimensional PESs explicitly depending on hundreds of degrees of freedom is required. This cannot be achieved in a brute-force approach by simply increasing the number of input nodes, because a systematic mapping of the associated configuration space is too costly. Further, the efficiency of the NN evaluation decreases with increasing number of input nodes. Finally, the internal structure of the NN constitutes a major obstacle in that the input nodes of the NN are "ordered". Each node in the input layer is connected to all nodes in the first hidden layer, but the numerical values of the connecting weights are all different. For larger systems containing many atoms there is necessarily an invariance of the total energy with respect to the exchange of atoms of the same element. Any high-dimensional NN approach must take this invariance into account. Recently, several extensions of the NN methodology have been suggested to address systems with a large number of degrees of freedom.

An extension of NN PESs to in principle arbitrary dimensionality has been proposed by employing the NN to fit the many-body term of an empirical potential[75]. Specifically, the functional form of the Tersoff potential has been used[15]:

$$V_{\mathrm{Tersoff}} = \frac{1}{2}\sum_i \sum_{j\neq i} f_c(r_{ij})\left[V_R(r_{ij} - b_{ij}V_A(r_{ij}))\right] \tag{29}$$

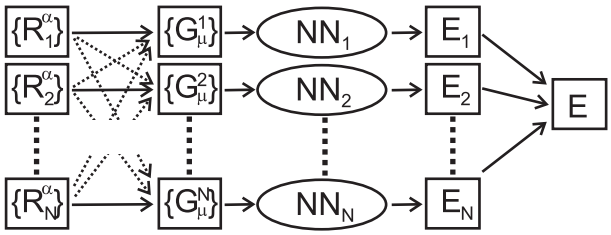


Figure 8. Schematic structure of a high-dimensional neural network potential for a condensed systems containing N atoms[79]. Each atom contributes the energy E_i to the total energy E. The local geometric environment of each atom with the Cartesian coordinates $\left\{R_i^\alpha\right\}$ is described by a set of symmetry functions $\left\{G_\mu^i\right\}$. The set of symmetry functions of each atom thus depends on the positions of all other atoms up to a cutoff radius as described by the dotted arrows. The symmetry functions are then used as input values for atomic neural networks (NN), which yield the E_i. The NN can be applied to any system size, because all atomic NNs have are identical, and for each additional atom one line has to be added to this scheme.

V_R and V_A are repulsive and attractive two-body terms. The attractive term is scaled by a many-body term b_{ij}, which is now replaced by a NN. The method has been applied to a binary system containing C and H[75] and to elemental silicon[76,77]. Its main drawbacks are the constrained functional form due to the combination with a rather simple empirical potential and the limitation to binary systems. Still, applications to large bulk structures and clusters with an explicit energy dependence on all atomic degrees of freedom have become possible showing promising results. Recently, this approach has been reformulated in terms of a general many-body expansion including two-, three- and four-body interactions[78].

Another approach for the construction of high-dimensional NN PESs avoiding any incorporation of empirical functional forms is based on a decomposition of the total energy E of the system into atomic energy contributions E_i[79]. This ansatz, which is also used in many empirical potentials, is based on the assumption that the energy contribution of each atom i in the system is determined by its local chemical environment up to a certain cutoff radius R_c. The total energy of the system is then constructed as a sum of the energy contributions of all atoms in the system

$$E = \sum_i E_i \tag{30}$$

The atomic energy contributions are calculated by atomic neural networks. The procedure is shown schematically in Fig. 8. Each atom in the system corresponds to one line in this scheme. First, the Cartesian coordinates $R_i = \{X_i, Y_i, Z_i\}$ of atom i are transformed to a set of symmetry functions $\{G_i\}$, which describe the local environment of this atom. The $\{G_i\}$ thus can be regarded as a kind of structural fingerprint, which is then used as input for an atomic NN. Typically 40-50 symmetry functions are used for each atom. This corresponds to an effective reduction of the dimensionality of the problem, because for each atom only neighbors inside the cutoff sphere determine the symmetry function values as indicated by the dotted arrows in Fig. 8. The cutoff radius R_c is typically chosen in the order of 6 Å, and the cutoff function f_c is defined as

$$f_{\mathrm{c}}\left(R_{ij}\right) = \begin{cases} 0.5 \cdot \left[\cos\left(\frac{\pi R_{ij}}{R_{\mathrm{c}}}\right) + 1\right] & \text{for } R_{ij} \leq R_{\mathrm{c}} \\ 0 & \text{for } R_{ij} > R_{\mathrm{c}}, \end{cases} \tag{31}$$

with R_{ij} being the distance between atom i and its neighbor j. At the cutoff radius f_c has zero value and slope. For the description of the local atomic environments many-body terms are used that depend explicitly on the positions of all atoms in the cutoff sphere. A "radial" symmetry function

$$G_i = \sum_j e^{-\eta_1 (R_{ij} - R_\mathrm{s})^2} \cdot f_\mathrm{c}\left(R_{ij}\right) \tag{32}$$

and an "angular" symmetry function for the angle $\theta = \frac{\mathbf{R_{ij}} \cdot \mathbf{R_{ik}}}{R_{ij} R_{ik}}$ centered at atom i, with $\mathbf{R_{ij}} = \mathbf{R_i} - \mathbf{R_j}$,

$$\begin{aligned} G_i = 2^{1-\zeta} \sum_\theta \left(1 + \gamma \cos\theta\right)^\zeta \cdot e^{-\eta_2 \left(R_{ij}^2 + R_{jk}^2 + R_{jk}^2\right)} \\ \cdot f_\mathrm{c}\left(R_{ij}\right) \cdot f_\mathrm{c}\left(R_{ik}\right) \cdot f_\mathrm{c}\left(R_{jk}\right) \end{aligned} \tag{33}$$

have been employed. These functions depend on parameters η_1, η_2, γ, R_s and ζ, which define the region of configuration space described by the symmetry functions, and which are not optimized. A set of many different symmetry functions of these types is typically used differing in the parameter values. The radial functions can be interpreted as effective coordination numbers at various distances, the angular functions as angle distributions. Further details on the symmetry functions can be found elsewhere[79,80].

The outputs of the atomic NNs are the atomic energies E_i, which are finally added to yield the total energy of the system. This scheme has several advantages: first, the topology as well as the weight parameters of all atomic NNs are identical, thus exchanging two atoms in the system does not change the total energy. Secondly, once the weight parameters of the atomic NN have been determined, the NN PES can be applied to systems of varying size, because for each additional atom simply an atomic NN is added in the scheme in Fig. 8. Finally, because all symmetry functions are high-dimensional many-body functions, they are very well suited to describe systems with strong many-body interactions like metals. A drawback of this approach is related to the finite range of the atomic interactions, which is defined by the cutoff radius. If long-range interactions are important, which is typically the case for systems with charge transfer, the accuracy of the approach will be strongly reduced unless these interactions are explicitly taken into account. Thus, up to now only a few applications for elemental solids exist[81,80,82].

6 Discussion

In general, neural networks provide a very general and unbiased way to construct accurate potential-energy surfaces. However, there are also some drawbacks that need to be discussed. First of all, NNs provide analytic energy expressions which do not allow for a physical interpretation of individual terms. Internally, NNs remain a "black box", and the reliability of the PES has to be checked carefully.

A serious problem of NN potentials is related to the very flexible functional form, which is the reason for the accurate representation of the training points, but is *a priori* non-physical. Consequently, neural networks are very accurate for the energy prediction of structures similar to the structures included in the training set, but they can spectacularly fail for very different structures. An illustrative example is given in Fig. 9. It shows the

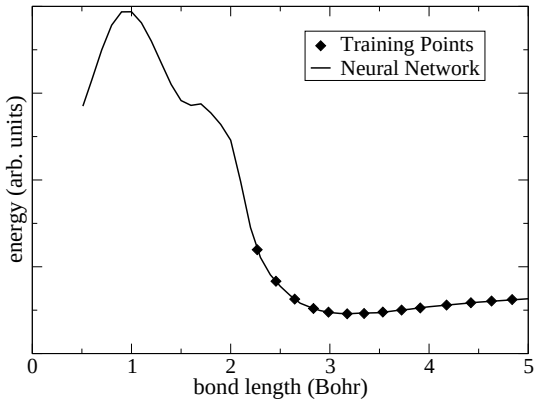


Figure 9. Demonstration of the poor extrapolation capabilities of neural networks (NN) for the binding curve of a dimer. In the range well represented by training points (solid diamonds) the NN potential (straight line) is very accurate, but because the NN does not include any assumptions on the functional form of the potential-energy surface, the highly repulsive part of the potential at short bond lengths is not described correctly. Instead the superposition of the tails of the activation functions give rise to an unphysical and unpredictable shape of the curve.

NN fit of a typical pair-potential of a dimer using a set of training points indicated as diamonds. In the range of bond lengths well-represented by the training points the NN energy curve is close to the reference data. For very short bonds however, the potential should be highly repulsive. Instead an obviously unphysical curve shape is obtained arising from a superposition of the tails of the activation functions. Because no training points are present in this range, this problem cannot be detected by an inspection of the RMSEs of the training set and the test set, which is typically distributed in the same range of values as the training set. However, it is straightforward to identify these regions by comparing for each input node of the NN the minimum and maximum values, which are present in the training set, with the symmetry functions values of a new atomic configuration. If the energy is requested for a structure whose symmetry function values are outside this range, the energy prediction should be taken with great care. Neural networks are designed for accurate interpolation, but they are not reliable in case of extrapolation. In molecular dynamics applications one has to make sure that the NN potential is "complete", i.e., there must not be any energetically accessible parts of the configuration space which are not well represented in the training set.

An advantage of NN potentials is that in contrast to classical force fields[3–7] they do not require a classification of atoms in terms of functional groups or hybridization state, and no bonds need to be specified. Neural network potentials are intrinsically "reactive" and in the course of an MD simulation based on a NN potential bonds can be broken or formed. Like in the underlying electronic structure calculations, just the atomic positions have to be provided in form of a suitable set of coordinates.

The use of system-specific symmetry functions allows to include information on the molecular symmetry in the NN, but often the construction of suitable symmetry functions is not straightforward. A combination of symmetry functions with a structural partitioning of the system into local environments as described in Section 5.3 now enables an application

to high-dimensional systems, but the exponential growth in the number of possible atomic configurations with the number of elements makes it unlikely for the near future that NN potentials will become a serious competitor for classical force fields e.g. in the field of biochemistry, unless their major advantages of "reactivity" is sacrificed by reducing the NN to fitting given parts of force fields. On the other hand, chemical problems involving a large number of atoms but a moderate number of elements, like in materials science, solid state chemistry, and some fields of surface science might strongly benefit from NN potentials in the future.

7 Concluding Remarks

In summary, the basic methodology and some technical aspects of the application of artificial neural networks to the construction of potential-energy surfaces for chemical reactions have been reviewed. Neural networks represent flexible functions, which are able to reproduce a given set of electronic structure data with high accuracy and to provide interpolated energies for similar structures. The resulting NN PES is computationally many orders of magnitude faster to evaluate than efficient electronic structure methods like density functional theory. The NN methodology itself is not bound to any particular reference total energy method and can also be applied in combination with wave-function based quantum chemical methods. The availability of analytic energy gradients enables a straightforward calculation of atomic forces for molecular dynamics simulations. The high flexibility, which is the reason for the numerical accuracy, is also the major drawback of NNs, namely the non-physical functional form. A large number of reference points is required to train a reliable NN PES and an extrapolation of the energy of structures very different from the structures included in the training set is not possible. The main future challenges are an extension of the applicability of NN potentials to high-dimensional multicomponent systems and the development of more systematic strategies for the construction and transferability checks of these potentials.

References

1. A. Szabo, and N.S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover, 1996.
2. W. Koch, and M.C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley-VCH, 2001.
3. N.L. Allinger, Y.H. Yuh, and J.-H. Lii, J. Am. Chem. Soc. **111**, 8551 (1989).
4. S.L. Mayo, B.D. Olafson, and W.A. Goddard III, J. Phys. Chem. **94**, 8897 (1990).
5. A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard III, and W.M. Skiff, J. Am. Chem. Soc. **114**, 10024 (1992).
6. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, J. Comp. Chem. **4**, 187 (1983).
7. W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, Jr., K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
8. J.E. Lennard-Jones, Proc. Phys. Soc. **43**, 461, 1931.

9. A.C.T. van Duin, A. Strachan, S. Stewman, Q. Zhang, X. Xu, and W.A. Goddard III, J. Phys. Chem. A **107**, 3803 (2003).
10. A. C. T. van Duin, S. Dasgupta, F. Lorant, W.A. Goddard III, J. Phys. Chem. A **105**, 9396 (2001).
11. A. Strachan, E.M. Kober, A.C.T. van Duin, J. Oxgaard, W.A. Goddard III, J. Chem. Phys. **122**, 054502 (2005).
12. K.D. Nielson, A.C.T. van Duin, J. Oxgaard, W.-Q. Deng, and W.A. Goddard III, J. Phys. Chem. A **109**, 493 (2005).
13. M.J. Buehler, A.C.T. van Duin, and W.A. Goddard III, Phys. Rev. Lett. **96**, 95505 (2006).
14. J. Tersoff, Phys. Rev. B **39**, 5566, 1989.
15. J. Tersoff, Phys. Rev. Lett. **56**, 632, 1986.
16. D.W. Brenner, and B.J. Garrison, Phys. Rev. B **34**, 1304, 1986.
17. F.H. Stillinger, and T.A.Weber, Phys. Rev. B **31**, 5262, 1985.
18. A. Groß, S. Wilk, and M.Scheffler, Phys. Rev. Lett. **75**, 2718, 1995.
19. G. Wiesenekker, G.J. Kroes, and E.J. Baerends J. Chem. Phys. **104**, 7344, 1996.
20. C.M. Wei, A. Groß, and M.Scheffler, Phys. Rev. B **57**, 15572, 1998.
21. C. Crespos, M.A. Collins, E. Pijper, and G.J. Kroes, Chem. Phys. Lett. **376**, 566, 2003.
22. C. Crespos, M.A. Collins, E. Pijper, and G.J. Kroes, J. Chem. Phys. **120**, 2392, 2004.
23. W. McCulloch, and W. Pitts, Bull. Math. Biophys. **5**, 115, 1943.
24. C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.
25. K. Hornik, M. Stinchcombe, and H. White, Neural Networks **2**, 359, 1989.
26. J. Zupan, and J. Gasteiger, *Neural Networks for Chemists*, VCH, Weinheim, 1993.
27. W. Duch, G.H.F. Diercksen, Comp. Phys. Comm. **82**, 91, 1994.
28. B.G. Sumpter, C. Getino, and D.W. Noid, Ann. Rev. Phys. Chem. **45**, 439, 1994.
29. J. Zupan, and J. Gasteiger, Anal. Chim. Acta **248**, 1, 1991.
30. J. Gasteiger, and J. Zupan, Angew. Chem. **105**, 510, 1993.
31. R. M. Agrawal, A. N. A. Samadh, L. M. Raff, M. T. Hagan, S. T. Bukkapatnam, and R. Komanduri J. Chem. Phys. **123**, 224711, 2005.
32. G. Reibnegger, G. Weiss, G. Werner-Felmayer, G. Judmaier, and H. Wächter, Proc. Natl. Acad. Sci. USA **88**, 11426, 1991.
33. M. Keil, R.E. Exner, J. Brickmann, J. Comp. Chem. **25**, 779, 2004.
34. G. Toth, N. Kiraly, and A. Vrabecz, J. Chem. Phys. **123**, 174109, 2005.
35. T.H. Fischer, W.P. Petersen, and H.P. Lüthi, J. Comp. Chem. **16**, 923, 1995.
36. J. Gasteiger, A. Teckentrup, L. Terfloth, and S. Spycher, J. Phys. Org. Chem. **16**, 232, 2003.
37. X. Zheng, L.H. Hu, X.J. Wang, and G.H. Chen, Chem. Phys. Lett. **390**, 186, 2004.
38. M. Sugawara, Comp. Phys. Comm. **140**, 366, 2001.
39. L.H. Holley, and M. Karplus, Proc. Natl. Acad. Sci. USA **86**, 152, 1988.
40. J.M. Boone, V.G. Sigillito, and G.S. Shaber, Med. Phys. **17**, 234, 1990.
41. F. Scarselli, and A.C. Tsoi, Neural Networks **11**, 15, 1998.
42. J.-G. Attali, and G. Pages, Neural Networks **10**, 1069, 1997.
43. J. Hertz, A. Krogh, and R.G. Palmer, *Introduction to the theory of neural computation*, Addison Wesley, Reading, 1996.

44. R. Rojas, *Theorie der Neuronalen Netze*, Springer, Heidelberg, 1996.
45. Y. LeCun, L. Bottou, G.B. Orr, and K.R. Müller, in *Neural Networks: Tricks of the Trade*, G.B. Orr, and K.R. Müller (eds.), Springer, Heidelberg, 1998.
46. K.B. Aspeslagh, Masters Thesis, Wheaton College, Norton, MA, USA, 2000.
47. T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, J. Chem. Phys. **103**, 4129, 1995.
48. C. Munoz-Caro, and A. Nino, Computers Chem. **22**, 355, 1998.
49. S. Lorenz, A. Groß, and M. Scheffler, Chem. Phys. Lett. **395**, 210, 2004.
50. H. Gassner, M. Probst, A. Lauenstein, and K. Hermansson, J. Phys. Chem. A **102**, 4596, 1998.
51. J. Behler, S. Lorenz, and K. Reuter, J. Chem. Phys. **127**, 014705, 2007.
52. A. J. Skinner, and J. Q. Broughton, Modelling Simul. Mater. Sci. Eng. **3**, 371, 1995.
53. R. Fletcher, and C.M. Reeves, Comput. J. **7**, 149, 1964.
54. E. Polak, and G. Ribiere, Revue Francaise Informat. Recherche Operationelle **16**, 35, 1969.
55. E. Polak, *Computational Methods in Optimization*, Academic Press, New York, 1971.
56. T.H. Blank, and S.D. Brown, J. Chemometrics **8**, 391, 1994.
57. S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, Science **220**, 671, 1983.
58. J. B. Witkoskie, and D. J. Doren, J. Chem. Theory Comput. **1**, 14, 2005.
59. A. Gelb, *Applied Optimal Estimation*, MIT Press, Cambridge, MA, USA, 1974.
60. F. V. Prudente, and J. J. S. Neto, Chem. Phys. Lett. **287**, 585, 1998.
61. A.C.P. Bittencourt, F.V. Prudente, and J.D.M. Vianna, Chem. Phys. **297**, 153, 2004.
62. F. V. Prudente, P. H. Acioli, and J. J. S. Neto, J. Chem. Phys. **109**, 8801, 1998.
63. T.M. Rocha Filho, Z.T. Oliveira, Jr., L.A.C. Malbouisson, R. Gargano, and J.J. Soares Neto, Int. J. Quant. Chem. **95**, 281, 2003.
64. D. F. R. Brown, M. N. Gibbs, and D. C. Clary, J. Chem. Phys. **105**, 7597, 1996.
65. K. T. No, B. H. Chang, S. Y. Kim, M. S. Jhon, and H. A. Scheraga, Chem. Phys. Lett. **271**, 152, 1997.
66. K.-H. Cho, K.T. No, and H.A. Scheraga, J. Mol. Struct. **641**, 77, 2002.
67. S. Manzhos, and T. Carrington, Jr., J. Chem. Phys. **125**, 084109, 2006.
68. S. Manzhos, and T. Carrington, Jr., J. Chem. Phys. **125**, 194105, 2006.
69. S. Manzhos, X. Wang, R. Dawes, and T. Carrington, Jr. J. Phys. Chem. A **110**, 5295, 2006.
70. S. Manzhos, and T. Carrington, Jr., J. Chem. Phys. **127**, 014103, 2007.
71. S. Lorenz, M. Scheffler, and A. Groß, Phys. Rev. B **73**, 115431, 2006.
72. J. Behler, K. Reuter, and M. Scheffler, Phys. Rev. B **77**, 115421, 2008.
73. J. Behler, B. Delley, S. Lorenz, K. Reuter, and M.Scheffler, Phys. Rev. Lett. **94**, 36104, 2005.
74. J. Ludwig, and D.G. Vlachos, J. Chem. Phys. **127**, 154716, 2007.
75. S. Hobday, R. Smith, and J. Belbruno, Modelling Simul. Mater. Sci. Eng. **7**, 397, 1999.
76. S. Hobday, R. Smith, and J. Belbruno, Nucl. Instr. Meth. Phys. Res. B **153**, 247, 1999.
77. M. Malshe, R. Narulkar, L.M. Raff, M. Hagan, S. Bukkapatnam, and R. Komanduri, J. Chem. Phys. **129**, 044111, 2008.
78. A. Bholoa, S.D. Kenny, and R. Smith, Nucl. Instr. Meth. Phys. Res. B **255**, 1, 2007.

79. J. Behler, and M. Parrinello, Phys. Rev. Lett. **98**, 146401, 2007.
80. J. Behler, R. Martoňák D. Donadio, and M. Parrinello, phys. stat. sol. (b) **245**, 2618, 2008.
81. J. Behler, R. Martoňák D. Donadio, and M. Parrinello, Phys. Rev. Lett. **100**, 185501, 2008.
82. H. Eshet, J. Behler, and M. Parrinello, in preparation (2009).

Multiscale Modelling of Magnetic Materials: From the Total Energy of the Homogeneous Electron Gas to the Curie Temperature of Ferromagnets

Phivos Mavropoulos

Institute for Solid State Research and Institute for Advanced Simulation
Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: Ph.Mavropoulos@fz-juelich.de

A widely used multiscale approach for the calculation of temperature-dependent magnetic properties of materials is presented. The approach is based on density functional theory, which, starting only from fundamental physical constants, provides the ground-state magnetic structure and a reasonable parametrization of the excited-state energies of magnetic systems, usually in terms of the Heisenberg model. Aided by statistical mechanical methods for the solution of the latter, the approach is at the end able to predict to within 10-20% high-temperature, material-specific magnetic properties such as the Curie temperature or the correlation function without the need for any fitting to experimental results.

1 Introduction

The physics of magnetism in materials spans many length scales. Starting from the formation of atomic moments by electron spins on the Ångström scale, it extends through the inter-atomic exchange interaction on the sub-nanometer scale to the formation of magnetic domains and hysteresis phenomena on the mesoscopic and macroscopic scale. In addition, the physics of magnetism spans many energy scales. The moments formation energy can be of the order of a few eV, the inter-atomic exchange of the order of 10-100 meV, elementary spin-wave excitations are of the order of 1-10 meV, while the magnetocrystalline anisotropy energy can be as low as a μeV. An energy-frequency correspondence implies the importance of as many time scales: from characteristic times of femto-seconds, related to the inter-atomic electron hopping and the atomic moments, through pico-seconds, related to the magnonic excitations, to seconds, hours or years related to the stability of a macroscopic magnetic configuration, e.g. defining a bit of information on a hard disc drive.

Clearly, a unified description of all these scales on the same footing is impossible. While many-body quantum mechanical calculations are necessary for the understanding of the small length scale phenomena, simple, possibly classical models have to suffice for the large scale. In this situation, multiscale modelling can provide a description on all scales, without adjusting parameters to experiment, but rather using results from one scale as input parameters to the model of the next scale. The scope of this manuscript is the presentation of such an approach, called here the Multiscale Programme, which is widely applied in present day calculations of magnetic material properties.

The manuscript is meant to serve as an introduction to the subject, not as a review. The list of references is definitely incomplete, reflecting only some suggested further reading. Finally, it should be noted that there are other multiscale concepts in magnetism, mainly in

the direction of micromagnetics and time evolution of the magnetization, as mentioned in Sec. 7.3. This type of multiscale modelling is an important field, however its description is beyond the scope of the present manuscript.

2 Outline of the Multiscale Programme

The outline of the Multiscale Programme can be summarized by the following steps, which will be explained in more detail in the next sections:

1. Calculation of the exchange-correlation energy of the electron gas, $E_{\rm xc}[\rho]$, as a functional of the electron density $\rho(\vec{r})$ by quantum Monte Carlo calculations and/or many-body theory.

2. Proper (approximate) parametrization of $E_{\rm xc}[\rho]$, usually in terms of ρ and $\nabla\rho$.

3. Use of $E_{\rm xc}[\rho]$ in density functional calculations for the unconstrained ground-state properties of a magnetic material (in particular, ground state atomic magnetic moments $\vec{M}_n$ and total energy $E^0_{\rm tot}$).

4. Use of $E_{\rm xc}[\rho]$ in *constrained* density functional calculations for the ground-state properties of a magnetic material under the influence an external, position-dependent magnetic field that forces a rotation of the magnetic moments $\{\vec{M}_n\}$, resulting in a total energy $E^{\rm constr}_{\rm tot}(\{\vec{M}_n\})$.

5. The adiabatic hypothesis: assumption that the time-scale of low-lying magnetic excitations is much longer than the one of inter-site electron hopping, so that $E^{\rm constr}_{\rm tot}(\{\vec{M}_n\})$ is a good approximation to the total energy of the excited state.

6. Correspondence to the Heisenberg hamiltonian under the assumption that $\Delta E(\{\vec{M}_n\}) := E^{\rm constr}_{\rm tot}(\{\vec{M}_n\}) - E^0_{\rm tot} \simeq -\frac{1}{2}\sum J_{nn'}\,\vec{M}_n\cdot\vec{M}_{n'} + {\rm const.}$

7. Solution of the Heisenberg hamiltonian $H = -\frac{1}{2}\sum_{nn'} J_{nn'}\,\vec{M}_n\cdot\vec{M}_{n'}$, e.g. for the Curie temperature, via a Monte Carlo method.

Steps 3 and 6 are connecting different models to each other.

3 Principles of Density Functional Theory

The most widely used theory for quantitative predictions with no adjustable parameters in condensed matter physics is density functional theory (DFT). “No adjustable parameters” means that, in principle, only fundamental constants are taken from experiment: the electron charge, Planck’s constant, and the speed of light in vacuum. Given these (and the types of atoms that are present in the material of interest), DFT allows to calculate ground-state properties of materials, such as the total energy, ground-state lattice structure, charge density, magnetization, etc. Naturally, since in practice the method relies on approximations to the exchange and correlation energy of the many-body electron system, the results are not always quantitatively or even qualitatively correct.

3.1 The Hohenberg-Kohn theorems and the Kohn-Sham ansatz

Density functional theory relies on the theorems of Hohenberg and Kohn.[1] Loosely put, these state that the ground-state wave function of a many-electron gas (under the influence of an external potential) is uniquely defined by the ground-state density (ground-state wavefunctions and densities are in one-to-one correspondence), and that an energy functional of the density exists that is stationary at the ground-state density giving the ground-state energy. Thus a variational scheme (introduced by Kohn and Sham[2]) allows for minimization of the energy functional in terms of the density, yielding the ground-state density and energy. Within the Kohn-Sham scheme[2] for this minimization, an auxiliary system of non-interacting (with each other) electrons is introduced, obeying a Schrödinger-like equation in an effective potential $V_{\rm eff}$. The effective potential includes the Hartree potential and exchange-correlation effects which depend explicitly on the density, as well as the "external" potential of the atomic nuclei (external in the sense that it does not arise from the electron gas). The Schrödinger-like equation must then be solved self-consistently so that the density is reproduced by the auxiliary electron system. In order for the scheme to work, a separation of the total energy functional is done:

$$E_{\rm DFT}[\rho] = T_{\rm n.i.}[\rho] + E_{\rm ext}[\rho] + E_{\rm H}[\rho] + E_{\rm xc}[\rho]. \tag{1}$$

Here, $T_{\rm n.i.}$ is the kinetic energy of the auxiliary non-interacting electrons, $E_{\rm ext} = \int d^3r\, \rho(\vec{r})\, V_{\rm ext}(\vec{r})$ is the energy due to the external potential (e.g., atomic nuclei), $E_{\rm H} = -e^2\frac{1}{2}\int d^3r \int d^3r' \rho(\vec{r})\rho(\vec{r}')/|\vec{r}-\vec{r}'|$ is the Hartree energy, and $E_{\rm xc}$ is "all that remains", i.e., the exchange and correlation energy. All but the latter can be calculated with arbitrary precision, while $E_{\rm xc}$ requires some (uncontrolled) approximation which also determines the accuracy of the method.

In practice, DFT calculations rely on a local density approximation (LDA) to the exchange-correlation energy. This means that $E_{\rm xc}[\rho]$ is approximated by $E_{\rm xc}^{\rm LDA}[\rho] = \int d^3r\, \varepsilon_{\rm xc}^{\rm hom}(\rho(\vec{r}))\, \rho(\vec{r})$, where $\varepsilon_{\rm xc}^{\rm hom}(\rho)$ is the exchange-correlation energy per particle for a homogeneous electron gas of density ρ. In case of spin-polarized calculations, the spin density $\vec{m}(\vec{r})$ must be included, and ρ is replaced by the density matrix $\boldsymbol{\rho}(\vec{r}) = \rho(\vec{r})\,\mathbf{1} + \vec{\boldsymbol{\sigma}}\cdot\vec{m}(\vec{r})$ ($\vec{\boldsymbol{\sigma}}$ are the Pauli matrices and $\mathbf{1}$ the unit matrix); then we have the local spin density approximation (LSDA). Gradient corrections, taking into account also $\nabla\boldsymbol{\rho}$, lead to the also widely used generalized gradient approximation (GGA).

Henceforth we will denote by $\rho_{\rm min}$, $\vec{m}_{\rm min}$, and $\boldsymbol{\rho}_{\rm min}$ the density, spin density, and density matrix that yield the minimum of energy functionals (either within DFT or constrained DFT, to be discussed in Sec. 5.1). These can be found by application of the Rayleigh-Ritz variational principle to eq. (1) which leads to the Schrödinger-like equation:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_{\rm eff}(\vec{r}) + \vec{\boldsymbol{\sigma}}\cdot\vec{B}_{\rm eff}(\vec{r}) - E_i\right)\begin{pmatrix}\psi_{i\uparrow}(\vec{r})\\ \psi_{i\downarrow}(\vec{r})\end{pmatrix} = 0. \tag{2}$$

This is the first of the Kohn-Sham equations for the one-particle eigenfunctions $\psi_{\uparrow,\downarrow}(\vec{r};E)$ (dependent on spin 'up' ($\uparrow$) or 'down' ($\downarrow$) with respect to a local magnetization direction $\hat{\mu}(\vec{r})$ along $\vec{B}_{\rm eff}(\vec{r})$) and eigenenergies E_i of the auxiliary non-interacting-electron system. The set of Kohn-Sham equations is completed by the expressions for charge and spin

density,

$$\rho(\vec{r}) = \sum_{E_i \leq E_\mathrm{F}} \left(|\psi_{i\uparrow}(\vec{r})|^2 + |\psi_{i\downarrow}(\vec{r})|^2 \right) \tag{3}$$

$$\vec{m}(\vec{r}) = \hat{\mu}(\vec{r}) \sum_{E_i \leq E_\mathrm{F}} \left(|\psi_{i\uparrow}(\vec{r})|^2 - |\psi_{i\downarrow}(\vec{r})|^2 \right), \tag{4}$$

and the requirement for charge conservation that determines the Fermi level E_F,

$$N = \int d^3r \rho(\vec{r}) = \int d^3r \sum_{E_i \leq E_\mathrm{F}} \left(|\psi_{i\uparrow}(\vec{r})|^2 + |\psi_{i\downarrow}(\vec{r})|^2 \right). \tag{5}$$

Expressions (2-5) form the set of non-linear equations to be solved self-consistently in any DFT calculation. The effective potential $V_\mathrm{eff}(\vec{r})$ and magnetic field $\vec{B}_\mathrm{eff}(\vec{r})$ follow from functional derivation of the total energy terms $E_\mathrm{ext}[\rho]$, $E_\mathrm{H}[\rho]$, and $E_\mathrm{xc}[\boldsymbol{\rho}]$ with respect to $\rho(\vec{r})$ and $\vec{m}(\vec{r})$. At the end of the self-consistency procedure one obtains the ground-state energy $E^0_\mathrm{tot} = E_\mathrm{DFT}[\boldsymbol{\rho}_\mathrm{min}]$.

In terms of the single-particle energies E_i, the total energy (1) can be split into the "single-particle" part E_sp and a "double-counting" E_dc part as

$$E_\mathrm{DFT} = E_\mathrm{sp} + E_\mathrm{dc} \tag{6}$$

with

$$E_\mathrm{sp} = \sum_{E_i \leq E_\mathrm{F}} E_i \tag{7}$$

$$E_\mathrm{dc} = -\int d^3r \left(\rho(\vec{r})\, V_\mathrm{eff}(\vec{r}) + \vec{m}(\vec{r}) \cdot \vec{B}_\mathrm{eff}(\vec{r}) \right) + E_\mathrm{H}[\rho] + E_\mathrm{ext}[\rho] + E_\mathrm{xc}[\boldsymbol{\rho}]. \tag{8}$$

3.2 Exchange-correlation energy of the homogeneous electron gas

The total energy of the homogeneous electron gas can be split, following the Kohn-Sham ansatz, in three parts (here there is no external potential): the kinetic energy of a system of non-interacting electrons, $T^\mathrm{hom}_\mathrm{n.i.}$, the Hartree energy $E^\mathrm{hom}_\mathrm{H}$, and the exchange-correlation energy $E^\mathrm{hom}_\mathrm{xc}$ which is, by definition, all that remains :

$$E^\mathrm{hom}_\mathrm{xc}[\boldsymbol{\rho}] = E^\mathrm{hom}_\mathrm{tot}[\boldsymbol{\rho}] - T^\mathrm{hom}_\mathrm{n.i.}[\boldsymbol{\rho}] - E^\mathrm{hom}_\mathrm{H}[\rho] \tag{9}$$

Given that $T^\mathrm{hom}_\mathrm{n.i.}[\boldsymbol{\rho}]$ and $E^\mathrm{hom}_\mathrm{H}[\rho]$ are straightforward to calculate, an approximation to $E^\mathrm{hom}_\mathrm{tot}[\boldsymbol{\rho}]$ yields an approximation to $E^\mathrm{hom}_\mathrm{xc}[\boldsymbol{\rho}]$.

Analytic approximations to $E^\mathrm{hom}_\mathrm{xc}[\rho]$ have proven successful. In particular, in the first paper to introduce the LSDA[3] von Barth and Hedin presented an analytic calculation of the exchange-correlation energy and potential, including a suitable parametrization. This result, with a slightly different parametrization, was successfully applied to the calculation of electronic properties of metals (including effects of spin polarization).[4] A more accurate calculation of $E^\mathrm{hom}_\mathrm{tot}[\boldsymbol{\rho}]$, based on a quantum Monte Carlo method, was given by Ceperley and Alder,[5] the exchange-correlation part of which was parametrized by Vosko, Wilk and Nusair.[6] This is the most commonly used parametrization of LSDA, although in practice there is little difference in calculated material properties among the three parametrizations of LSDA.[3,4,6] Larger differences, usually towards increased accuracy, are provided when density gradient correction are included within the generalized gradient approximation (GGA).[7]

4 Magnetic Excitations and the Adiabatic Approximation

Density functional calculations reproduce, in many cases with remarkable accuracy, the ground-state magnetic moments of elemental or alloyed systems. Transition-metal ferromagnets (Fe, Co, Ni) and ferromagnetic metallic alloys (e.g. Heusler alloys, such as NiMnSb or Co_2MnSi), magnetic surfaces and interfaces are among the systems that are rather well described within the LSDA or GGA (with an accuracy of a few percent in the magnetic moment). On the other hand, materials where strong correlations play an important role, such as f-electron systems or antiferromagnetic transition metal oxides are not properly described within the LSDA or GGA, but in many cases corrections can be made by including additional semi-empirical terms in the energy and potential (as in the LSDA+U scheme).[8] As an example of the accuracy of the LSDA in the magnetic properties of transition metal alloys, fig. 1 shows experimental and theoretical results on the magnetic moments of Iron-, Cobalt-, and Nickel-based alloys.[9]

However, density functional theory is, in principle, a ground-state theory—at least in its usual, practical implementation. This means that the various approximations to the exchange-correlation potential, when applied, yield approximate values of ground-state energy, charge-density, magnetization, etc. Nevertheless, physical arguments can be used to derive also properties of excited states from DFT calculations. A basis for this is the *adiabatic approximation* (or *adiabatic hypothesis*), i.e., that the energies of some excitations, governed by characteristic frequencies much smaller than the ones of intra- and inter-site electron hopping, can be approximated by ground-state calculations. The adiabatic hypothesis is most often used in the calculation of phonon spectra, ab-initio molecular dynamics, or magnetic excitations.

In magnetic materials, two types of magnetic excitations can be distinguished: (i) the Stoner-type, or longitudinal, where the absolute value of the atomic moments changes, and (ii) the Heisenberg-type, or transverse, where the relative direction of the moments changes. Longitudinal excitations usually require high energies, of the order of the intra-atomic exchange (order of 1 eV); clearly this energy scale is far beyond the Curie temperature (ferromagnetic fcc Cobalt has the highest known Curie temperature at 1403 K, while 1 eV corresponds to 11605 K). Transverse excitations (magnons), on the other hand, are one or two orders of magnitude weaker, and are responsible for the ferromagnetic-paramagnetic phase transition.

The characteristic time scale of magnons is of the order of 10^{-12} seconds. On the other hand, inter-atomic electron hopping takes place in timescales of the order of 10^{-15} seconds. As a result, during the time that it takes a magnon to traverse a part of the system, it is expected that locally the electron gas has time to adjust and relax to a new ground state, defined by a constrained, position-dependent magnetization direction. This is the adiabatic hypothesis. For practical calculations, this means that the magnon energy can be found by using an additional position-dependent magnetic field to constrain the magnetic configuration to a magnon-like form (a so-called spin spiral), and calculating the resulting total energy. It should be noted here that the magnon energy arises from the change in electron inter-site hopping energy.

Essentially, the adiabatic hypothesis directs us to approximate the excited-state energy of one system (e.g., a ferromagnet) by the ground-state energy of a different system (a ferromagnet under the influence of constraining fields).

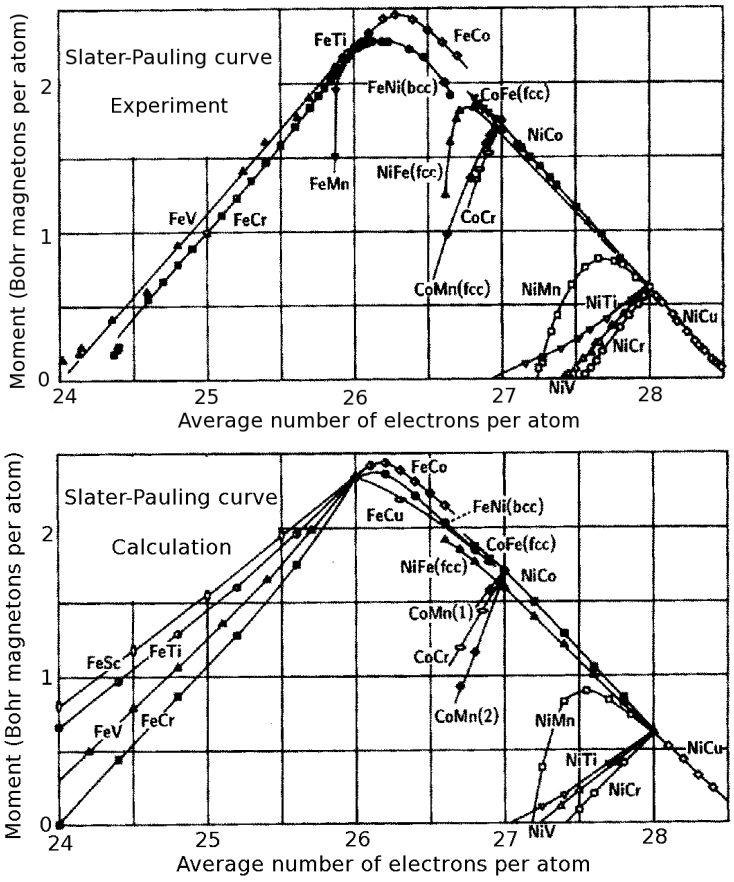


Figure 1. Magnetic moments of Fe, Co and Ni based transition-metal alloys, taken from Dederichs et al.[9] The theoretical results were calculated within the LSDA, using the Korringa-Kohn-Rostoker Green function method and the coherent potential approximation for the description of chemical disorder. The magnetization as a function of average number of electrons per atom follows the *Slater-Pauling rule*.

5 Calculations within the Adiabatic Hypothesis

In this section we discuss how the adiabatic hypothesis can be practically used to extract excited state energies from density functional calculations. The accuracy of the method is such that small energy differences, of the order of meV, can be reliably extracted from total energies of the order of thousands of eV; for instance, for fcc Co the calculated total energy per atom is approximately 38000 eV, while the nearest-neighbour exchange coupling is approximately 14 meV. Such accuracy is crucial for the success of the Multiscale Programme.

5.1 Constrained density functional theory for magnetic systems

Constrained DFT[10] includes an additional term to the energy functional, so that the system is forced to a specific configuration. For the case of interest here, the following functional must be minimized in order to obtain a particular configuration of magnetic moments $\{\vec{M}_n\}$:

$$E_{\mathrm{CDFT}}[\boldsymbol{\rho}; \{\vec{M}_n\}] = E_{\mathrm{DFT}}[\boldsymbol{\rho}] - \sum_n \int_{\text{Cell n}} d^3r\, \vec{H}_n \cdot \left[\vec{m}(\vec{r}) - \vec{M}_n\right]. \tag{10}$$

In this expression, $E_{\mathrm{DFT}}[\boldsymbol{\rho}]$ is the DFT energy functional (1) (e.g., in the LSDA or GGA), while the quantities $\{\vec{H}_n\}$ are Lagrange multipliers, physically interpreted as external magnetic fields acting in the atomic cells $\{n\}$; for convenience in notation we define $\vec{H}_n$ to be constant in the atomic cell n and zero outside. Furthermore, $\vec{m}(\vec{r})$ is the spin density, while $\vec{M}_n$ is the desired magnetic moment. Application of the Raleygh-Ritz variational principle to eq. (10) leads to the Schrödinger-like equation:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_{\mathrm{eff}}(\vec{r}) + \vec{\boldsymbol{\sigma}} \cdot \vec{B}_{\mathrm{eff}}(\vec{r}) + \vec{\boldsymbol{\sigma}} \cdot \sum_n \vec{H}_n - E_i\right) \begin{pmatrix} \psi_{i\uparrow}(\vec{r}) \\ \psi_{i\downarrow}(\vec{r}) \end{pmatrix} = 0. \tag{11}$$

This is just the Kohn-Sham equation (2) with an additional term containing the Lagrange multipliers $\vec{H}_n$ which act as an external Zeeman magnetic field (note that this is not really a magnetic field, in the sense that it is not associated to a vector potential, Landau levels, etc.). In practice, $\vec{H}_n$ is specified and the corresponding value of $\vec{M}_n$ is an output of the self-consistent calculation, calculated from the spin density as

$$\vec{M}_n = \int_{\text{Cell n}} d^3r\, \vec{m}(\vec{r}). \tag{12}$$

If a particular value of $\vec{M}_n$ is to be reached, then $\vec{H}_n$ has to be changed and $\vec{M}_n$ recalculated, until $\vec{M}_n$ reaches the pre-defined value. At the end the energy-functional minimization yields the density $\boldsymbol{\rho}_{\min}$, obeying the condition (12). Since the multipliers $\{\vec{H}_n\}$ enter equation (11) as external parameters, it is evident that the minimizing density $\boldsymbol{\rho}_{\min}$ and the constrained ground-state energy $E_{\mathrm{CDFT}}[\boldsymbol{\rho}_{\min}; \{\vec{M}_n\}]$ are functions of $\{\vec{H}_n\}$. Therefore, to simplify the notation when referring to the constrained ground state, we write $\boldsymbol{\rho}_{\min} = \boldsymbol{\rho}_{\min}[\{\vec{H}_n\}]$, $E_{\mathrm{CDFT}}[\boldsymbol{\rho}_{\min}; \{\vec{M}_n\}] = E_{\mathrm{CDFT}}[\{\vec{H}_n\}]$. Similarly, the multipliers $\{\vec{H}_n\}$ are functions of the constrained ground-state moments, and vice versa: $\vec{H}_n = \vec{H}_n[\{\vec{M}_m\}]$, $\vec{M}_n = \vec{M}_n[\{\vec{H}_m\}]$.

The total energy of the constrained state is given by

$$E_{\mathrm{tot}}^{\mathrm{constr}}(\{\vec{M}_n\}) := E_{\mathrm{CDFT}}[\{\vec{H}_n\}] = E_{\mathrm{DFT}}\left[\boldsymbol{\rho}_{\min}[\{\vec{H}_n\}]\right] \tag{13}$$

(the latter step, where the constrained ground-state density $\boldsymbol{\rho}_{\min}[\{\vec{H}_n\}]$ is taken as argument of the unconstrained density functional E_{DFT}, follows because the last part of eq. (10) vanishes for the self-consistent solution). In order to extract the excited state energy from eq. (13), a subtraction of the unconstrained-state energy from the constrained one is needed:

$$\Delta E[\{\vec{M}_n\}] = E_{\mathrm{CDFT}}[\{\vec{H}_n\}] - E_{\mathrm{CDFT}}[\{\vec{H}_n\} = 0] \tag{14}$$

$$= E_{\mathrm{tot}}^{\mathrm{constr}}[\{\vec{M}_n\}] - E_{\mathrm{tot}}^{0} \tag{15}$$

This can be susceptible to numerical errors, as the total energies are large quantities compared to the change in magnetization energy. There is an alternative to that route.[10,11] By taking advantage of the Helmann-Feynman theorem,

$$\frac{\partial E_{\mathrm{CDFT}}[\boldsymbol{\rho}_{\mathrm{min}};\{\vec{M}_m\}]}{\partial \vec{M}_n} = \vec{H}_n, \tag{16}$$

which rests on the variational nature of the energy around $\boldsymbol{\rho}_{\mathrm{min}}$, the energy difference can be calculated by an integration along a path from the ground-state moments $\vec{M}_n^{\mathrm{GS}} = \vec{M}_n[\{\vec{H}_m\} = 0]$ to the constrained end-state moments $\vec{M}_n$. Along this path, the Lagrange multipliers $\vec{H}_n[\{\vec{M}_m\}]$ are found by minimization of the constrained energy functional. We have:

$$E_{\mathrm{CDFT}}[\{\vec{H}_n\}] - E_{\mathrm{CDFT}}[\{\vec{H}_n\} = 0] = \sum_n \int_{\vec{M}_n^{\mathrm{GS}}}^{\vec{M}_n} d\vec{M}'_n \cdot \vec{H}_n[\{\vec{M}'_m\}]. \tag{17}$$

It should be noted, however, that this method can be numerically more expensive, as a number of self-consistent calculations are necessary along the path in order to obtain an accurate integration. In practice, the former method of total energy subtraction usually works rather well as long as care is taken for good spin-density convergence in the self-consistent cycle.

5.2 Magnetic force theorem

In principle, to find the excited-state energy $E^{\mathrm{tot}}_{\mathrm{constr}}[\{\vec{M}_n\}]$ one must perform a self-consistent calculation for the particular moments configuration $\{\vec{M}_n\}$. This can be computationally expensive. Fortunately, under certain conditions additional self-consistent calculations can be avoided by virtue of the *force theorem*.[12,13] This states that, under sufficiently small perturbations of the (spin) density, the total energy difference can be approximated by the difference of the occupied single-particle state energies, given by (7). As a consequence, for the total energy difference between the magnetic ground state and the magnetic state characterized by rotated moments $\{\vec{M}_n\}$, one has merely to perform a position-dependent rotation of the ground-state spin density $\vec{m}(\vec{r})$ to a new spin density $\vec{m}'(\vec{r})$ at each atom so that eq. (12) is satisfied, calculate the single-particle energies sum at this non-self-consistent spin density, and subtract the single-particle energies sum of the ground state:

$$\Delta E[\{\vec{M}_n\}] \simeq E_{\mathrm{sp}}[\rho, \vec{m}'] - E_{\mathrm{sp}}[\rho, \vec{m}]. \tag{18}$$

The calculation of $E_{\mathrm{sp}} = \sum_{E_i \leq E_{\mathrm{F}}} E_i$ requires the solution of eq. (11) (or eq. (2)), where the potentials V_{eff} and $\vec{B}_{\mathrm{eff}}$ enter explicitly instead of the densities ρ and $\vec{m}$. Therefore, in practice, the magnetic exchange-correlation potentials $\vec{B}_{\mathrm{eff}}$ are rotated for the energy estimation in eq. (18), instead of the spin density $\vec{m}$.

5.3 Reciprocal space analysis: generalized Bloch theorem

The elementary, transverse magnetic excitations in ferromagnetic crystals have, in a semi-classical picture, the form of spin spirals of wave-vector $\vec{q}$. If the ground-state magnetization M_0 is oriented along the z-axis, then in the presence of a spin spiral the spin density

and the exchange-correlation potential at the atomic cell at lattice point $\vec{R}_n$ are given in terms of a position-dependent angle $\phi_n = \vec{q}\cdot\vec{R}_n$ and an azimuthal angle θ:

$$\vec{m}(\vec{r}+\vec{R}_n) = m_0(\vec{r})\left(\sin\theta\,\cos(\vec{q}\cdot\vec{R}_n)\,\hat{x} + \sin\theta\,\cos(\vec{q}\cdot\vec{R}_n)\,\hat{y} + \cos\theta\,\hat{z}\right) \quad (19)$$

$$\vec{B}(\vec{r}+\vec{R}_n) = B_0(\vec{r})\left(\sin\theta\,\cos(\vec{q}\cdot\vec{R}_n)\,\hat{x} + \sin\theta\,\cos(\vec{q}\cdot\vec{R}_n)\,\hat{y} + \cos\theta\,\hat{z}\right) \quad (20)$$

This implies that the potential has a periodicity of the order of $1/q$, thus, for small q, the unit cell contains too many atoms to handle computationally. However, there is a *generalized Bloch theorem*,[14] by virtue of which the calculation can be confined to the primitive unit cell. The generalized Bloch theorem is valid under the assumption that the hamiltonian $\mathcal{H}$ (or equivalently the potential) obeys the transformation rule

$$\mathcal{H}(\vec{r}+\vec{R}_n) = \mathbf{U}(\vec{q}\cdot\vec{R}_n)\mathcal{H}(\vec{r})\mathbf{U}^{\dagger}(\vec{q}\cdot\vec{R}_n). \quad (21)$$

with the spin transformation matrix $\mathbf{U}$ defined by

$$\mathbf{U}(\vec{q}\cdot\vec{R}^n) = \begin{pmatrix} e^{-i\vec{q}\cdot\vec{R}_n/2} & 0 \\ 0 & e^{i\vec{q}\cdot\vec{R}_n/2} \end{pmatrix}. \quad (22)$$

This is true if the exchange-correlation potential has the form (20) and if the spin orbit coupling can be neglected. This transformation rule in spin space has as a consequence that the hamiltonian remains invariant under a *generalized translation* $\mathcal{T}_n = T_n\,\mathbf{U}(\vec{q}\cdot\vec{R}^n)$ which combines a translation in real space by the lattice vector $\vec{R}_n$, T_n, with a rotation in spin space, $\mathbf{U}(\vec{q}\cdot\vec{R}_n)$:

$$\mathcal{T}_n\,\mathcal{H}\,\mathcal{T}_n^{-1} = \mathcal{H}. \quad (23)$$

As a result of this invariance, using manipulations analogous to the ones that lead to the well-known Bloch theorem it can be shown that the spinor eigenfunctions are of the form

$$\psi_{\vec{k}}(\vec{r}) = e^{i\vec{k}\cdot\vec{r}}\begin{pmatrix} e^{-i\vec{q}\cdot\vec{r}}\,\alpha_{\vec{k}}(\vec{r}) \\ e^{+i\vec{q}\cdot\vec{r}}\,\beta_{\vec{k}}(\vec{r}) \end{pmatrix} \quad (24)$$

where $\alpha_{\vec{k}}(\vec{r})$ and $\beta_{\vec{k}}(\vec{r})$ are lattice-periodic functions, $\alpha_{\vec{k}}(\vec{r}+\vec{R}_n) = \alpha_{\vec{k}}(\vec{r})$ and $\beta_{\vec{k}}(\vec{r}+\vec{R}_n) = \beta_{\vec{k}}(\vec{r})$. In this way, given a particular spin-spiral vector $\vec{q}$, the calculation is confined in the primitive cell in real space (and in the first Brillouin zone in k-space) and is thus made computationally tractable.

In case that the atomic magnetic moments do not change appreciably under rotation, the energy differences $\Delta E(\vec{q};\theta)$ can be Fourier-transformed[15] in order to find the real-space excitation energies $\Delta E[\{\vec{M}_n\}]$. This is usually true when θ is small. Under this condition, the force theorem is also applicable, so that non-self-consistent calculations are sufficient to find the dispersion relation $\Delta E(\vec{q};\theta)$ for $\vec{q}$ in the Brillouin zone.

5.4 Real space analysis: Green functions and the method of infinitesimal rotations

For perturbations that are confined in space, the Green function method is most appropriate for the calculation of total energies. The reason is that it makes use of the Dyson equation for the derivation of the Green function of the perturbed system from the Green function of the unperturbed system, with the correct open boundary conditions taken into account

automatically. As opposed to this, in wave function methods for localized perturbations a solution of the Scrödinger (or Kohn-Sham) equation requires explicit knowledge of the boundary condition and a complicated coupling procedure in order to achieve continuity of the wavefunction and its first derivative at the boundary.

The Green function $\boldsymbol{G}(\vec{r},\vec{r}\,';E)$ corresponding to the Kohn-Sham hamiltonian of eq. (2) is a 2×2 matrix in spin space that obeys the equation

$$\left(-\tfrac{\hbar^2}{2m}\nabla^2+V_{\rm eff}(\vec{r})+\vec{\boldsymbol{\sigma}}\cdot\vec{B}_{\rm eff}(\vec{r})-E\right)\begin{pmatrix}G_{\uparrow\uparrow}(\vec{r},\vec{r}\,';E)\;G_{\uparrow\downarrow}(\vec{r},\vec{r}\,';E)\\ G_{\downarrow\uparrow}(\vec{r},\vec{r}\,';E)\;G_{\downarrow\downarrow}(\vec{r},\vec{r}\,';E)\end{pmatrix} = -\begin{pmatrix}1\;0\\0\;1\end{pmatrix}\delta(\vec{r}-\vec{r}\,'). \tag{25}$$

The particle density and spin density can be readily calculated from $\boldsymbol{G}$ as

$$\rho(\vec{r}) = -\frac{1}{\pi}{\rm Im}\int^{E_{\rm F}} dE\,{\rm Tr}_s\,\boldsymbol{G}(\vec{r},\vec{r}\,';E) \tag{26}$$

$$\vec{m}(\vec{r}) = -\frac{1}{\pi}{\rm Im}\int^{E_{\rm F}} dE\,{\rm Tr}_s\,[\vec{\boldsymbol{\sigma}}\,\boldsymbol{G}(\vec{r},\vec{r}\,';E)] \tag{27}$$

where ${\rm Tr}_s$ indicates a trace over spins. More generally, the Green function corresponding to a hamiltonian $\mathcal{H}$ obeys the equation $(E-\mathcal{H})\,\mathcal{G}(E)=1$. In case of a perturbation $\Delta\mathcal{V}$ to a hamiltonian $\mathcal{H}_0$, the Green function $\mathcal{G}(E)=(E-\mathcal{H})^{-1}$ to the new hamiltonian, $\mathcal{H}=\mathcal{H}_0+\Delta\mathcal{V}$, is related to the initial Green function, $\mathcal{G}_0(E)=(E-\mathcal{H}_0)^{-1}$, via the Dyson equation $\mathcal{G}(E)=\mathcal{G}_0(E)\,[1-\Delta\mathcal{V}\,\mathcal{G}_0(E)]^{-1}$. In practice, the latter equation is very convenient to use because it requires a minimal basis set. With some reformulation the Dyson equation forms the basis of the Korringa-Kohn-Rostoker (KKR) Green function method for the calculation of the electronic structure of solids[16] and impurities in solids.[17] Within the KKR method, the Green function is expanded in terms of regular ($R^n_{s;L}(\vec{r};E)$) and irregular ($H^n_{s;L}(\vec{r};E)$) scattering solutions of the Schrödinger equation for the atomic potentials embedded in free space. The index n denotes the atom, $L=(l,m)$ stands for a combined index for the angular momentum quantum numbers of an incident spherical wave, and s is the spin ($\uparrow$ or $\downarrow$). For a ferromagnetic system, where only spin-diagonal elements of the Green function exist, $G_{ss'}=G_s\delta_{ss'}$ in (25), the expansion reads:

$$\begin{aligned}G_s(\vec{r}+\vec{R}_n,\vec{r}\,'+\vec{R}_{n'};E) = -i\sqrt{\frac{2mE}{\hbar^2}}\sum_L R^n_{s;L}(\vec{r};E)H^n_{s;L}(\vec{r}\,';E)\,\delta_{nn'}\\ +\sum_{LL'}R^n_{s;L}(\vec{r};E)\,G^{nn'}_{s;LL'}(E)\,R^{n'}_{s;L'}(\vec{r}\,';E)\end{aligned} \tag{28}$$

for $|\vec{r}|<|\vec{r}\,'|$ (for $|\vec{r}|>|\vec{r}\,'|$, $\vec{r}$ and $\vec{r}\,'$ should be interchanged in the first term of the RHS). The coefficients $G^{nn'}_{s;LL'}(E)$ are called structural Green functions and are related to the structural Green functions of a reference system (e.g., free space) via an algebraic Dyson equation[16,17] which involves the spin-dependent scattering matrices $t^n_{s;LL'}(E)$. In case of a non-collinear magnetic perturbation in a ferromagnetic system, the method can be generalized in a straightforward way[13,18] yielding the total energy of the state, $E[\{\vec{M}_n\}]$. However, in the limit of infinitesimal rotations of the moments $\{\vec{M}_n\}$, perturbation theory can be employed in order to find the change in the density of states, and by application

of the force theorem, the change in total energy. Of particular interest for our discussion below is the result for the total energy change in second order when two moments $\vec{M}_n$ and $\vec{M}_{n'}$ are infinitesimally rotated:[19]

$$\frac{\delta^2 E}{\delta\vec{M}_n \delta\vec{M}_{n'}} = -\frac{1}{8\pi\,|\vec{M}_n|\,|\vec{M}_{n'}|}\mathrm{Im}\int^{E_\mathrm{F}} dE\,\mathrm{Tr}_L\left[\boldsymbol{G}_\uparrow^{nn'}\,(\boldsymbol{t}_\uparrow^{n'}-\boldsymbol{t}_\downarrow^{n'})\,\boldsymbol{G}_\uparrow^{n'n}\,(\boldsymbol{t}_\uparrow^{n}-\boldsymbol{t}_\downarrow^{n})\right] \tag{29}$$

In this formula, $\boldsymbol{G}_s^{nn'}(E)$ is the structural Green function of spin s in form of a matrix in L, L', while $\boldsymbol{t}_s^n(E)$ are again the scattering matrices. Tr_L denotes a trace in angular momentum quantum numbers. The derivatives on the LHS are implied to be taken only with respect to the angles of $\vec{M}_n, \vec{M}_{n'}$, not the magnitude.

6 Correspondence to the Heisenberg Model

The next step of the Multiscale Programme is to establish a correspondence between the density functional results and the parameters of a phenomenological model hamiltonian for magnetism. Usually, the classical Heisenberg model is used in order to derive the magnetism-related statistics up to (and even beyond) the Curie temperature, and we will focus on this. However, other models can be used for different purposes, such as the continuum model for micromagnetic or magnetization dynamics calculations. Also, even on the atomic scale, it is sometimes necessary to extend the Heisenberg model to non-rigid spins.

The classical Heisenberg hamiltonian for a system of magnetic moments $\{\vec{M}_n\}$ is

$$\mathcal{H} = -\frac{1}{2}\sum_{nn'} J_{nn'}\vec{M}_n\vec{M}_{n'}. \tag{30}$$

The quantities $J_{nn'}$ are called pair exchange constants, and they are assumed to be symmetric ($J_{nn'} = J_{n'n}$), while, by convention, $J_{nn} = 0$. The prefactor $1/2$ takes care of double-counting. The exchange constants fall off sufficiently fast with distance, so that only a finite amount of neighbours n' has to be considered in the sum for each n. Physically, it is well known that the exchange interaction results from the change of the electronic energy under rotation of the moments, not from the dipole-dipole interaction of the moments.

A correspondence to density functional calculations can be made due to the observation that

$$J_{nn'} = -\frac{\partial^2\mathcal{H}}{\partial\vec{M}_n\partial\vec{M}_{n'}} \tag{31}$$

assuming that, to a good approximation, the constrained DFT energy can be expanded to lowest order in the moments' angles as $E_\mathrm{tot}^\mathrm{constr}[\{\vec{M}_n\}] - E_\mathrm{tot}^0 = -\frac{1}{2}\sum_{nn'} J_{nn'}\vec{M}_n\vec{M}_{n'} +$ const. By computing $E[\{\vec{M}_n\}]$ within constrained DFT, the RHS can be evaluated, and $J_{nn'}$ can be found. Thus, the step from DFT to the Heisenberg model relies on accepting the equivalence of the DFT and Heisenberg-model excitation energies. As an example we see in fig. 2 the exchange constants of fcc Cobalt as a function of distance.

Additional terms to the Heisenberg hamiltonian (30) can also be evaluated in a similar way. For instance, the magnetocrystalline anisotropy energy is phenomenologically

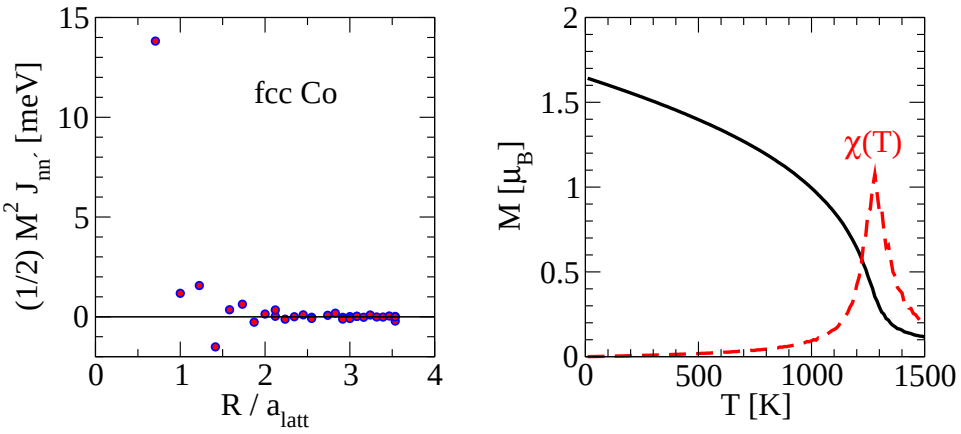


Figure 2. Left: Exchange constants as a function of inter-atomic distance in fcc Co calculated by the method of infinitesimal rotations. Right: Magnetization (in $\mu_{\rm B}$ per atom) and susceptibility χ as functions of temperature, calculated by a Monte Carlo method using the exchange constants of the left panel. The peak of susceptibility signals the Curie temperature. In the simulation a supercell of 1728 atoms was used. The experimentally found Curie temperature is 1403 K.

described by adding the term $-K\sum_n(\vec{M}_n\cdot\hat{\zeta})^2 = -KM\sum_n\cos^2\gamma_n$, where $\hat{\zeta}$ is a unit vector, usually along a high-symmetry crystal axis, and γ_n is the angle of the magnetic moment to this axis. The magnetocrystalline anisotropy, which stems from the spin-orbit coupling, induces the preference of a particular direction for the magnetic moments ($\pm\hat{\zeta}$), if $K > 0$, or in the plane perpendicular to this direction, if $K < 0$. By observing that $K = \frac{1}{2M}\partial^2\mathcal{H}/\partial\gamma^2|_{\gamma=0}$, the constant K can be harvested by fitting DFT total-energy results to the second derivative $\partial^2 E_{\rm CDFT}[\{\vec{M}_n\}]/\partial\gamma^2|_{\gamma=0}$. Furthermore, in all cases the validity of the phenomenological model can also be subjected to verification by DFT calculations.

Having established the correspondence to the Heisenberg model, there are two practical, widely used ways to calculate the exchange constants $J_{nn'}$. The first, used within Green function methods (KKR or linearized muffin-tin orbital (LMTO) Green function), is a direct combination of eqs. (29) and (31). The second, used within hamiltonian methods, is a Fourier transform of the spin-spiral energy $\Delta E(\vec{q};\theta)$.[15] It should be noted, however, that the assumption of rigid magnetic moment magnitudes, inherent in the Heisenberg model, is only an approximation. When the moment angles between nearest-neighbour atoms become large, the moments can change and the Heisenberg hamiltonian is not any more valid. The extent of this strongly depends on the material, as has been found by DFT calculations; therefore, the Heisenberg hamiltonian should only be considered as the lowest order expansion in the moments.

According to these considerations, the method of infinitesimal rotations should be ideal for calculating the $J_{nn'}$, while the Fourier transform of $\Delta E(\vec{q};\theta)$ is accurate only when θ is chosen small enough. However, this is not the whole story. At high temperatures, close to the Curie temperature, neighbouring moments can have larger respective angles, perhaps of the order of 30 degrees or more. Therefore some sort of "intelligent averaging" over

angles is called for, in order to increase the accuracy of results. The method of infinitesimal rotations can be systematically amended in this direction, as was proposed by Bruno,[20] while for the Fourier-transform method larger angles θ (perhaps of the order of 30 degrees) should be considered. We will return to this discussion in Section 8. We should also mention that the formalism, as it is presented in the present manuscript, neglects the orbital moments and their interaction. Such effects can become important especially for rare earths and actinides, which are, however, not well-described by local density functional theory due to the strong electron correlations in these systems.

7 Solution of the Heisenberg Model

Having established the correspondence between DFT results and the Heisenberg hamiltonian, and having identified the model parameters, a statistical-mechanical method is used in order to solve the Heisenberg model, if one is interested in thermodynamic properties, or a dynamical method is used if one is interested in time-dependent properties. In the former case, the Monte Carlo method, mean-field theory, and the random phase approximation (RPA) are most commonly used. For time-dependent properties we give a brief introduction to Landau-Lifshitz spin dynamics.

7.1 Thermodynamic properties and the Curie temperature

The Monte Carlo method is a stochastic approach to the solution of the Heisenberg model (and of course to many other problems in physics). It is based on a random walk in the configuration space of values of $\{\vec{M}_n\}$, but with an intelligently chosen probability for transition from each state to the next. The random walk must fulfill two requirements: (i) it must be ergodic, i.e., each point of the configuration space must be in principle accessible during the walk, and (ii) the transition probability between states A and B, $t_{A\to B}$, must obey the detailed balance condition, i.e., $P(A)\,t_{A\to B} = P(B)\,t_{B\to A}$, where $P(X) = \exp(-E(X)/k_{\mathrm{B}}T)$ is the Boltzmann probability for appearance of state X at temperature T, with $E(X)$ the energy of the state and k_{B} the Boltzmann constant. As long as these requirements are fulfilled, $t_{A\to B}$ is to be chosen in a way that optimizes the efficiency of the method. The most simple and widely-used way is the Metropolis algorithm,[21] in which $t_{A\to B} = P(B)/P(A) = \exp[(E(A) - E(B))/k_{\mathrm{B}}T]$ is taken for $E(A) < E(B)$ and $t_{A\to B} = 1$ otherwise. For further reading on the Monte Carlo method we refer the reader to the book by Landau and Binder.[22]

Within the Monte Carlo method, a simulation supercell is considered, which contains many atomic sites (e.g., $10 \times 10 \times 10$ for simulating a three-dimensional cubic ferromagnetic lattice). At each site, a magnetic moment $\vec{M}_n$ is placed, subject to interactions $J_{nn'}$ with the neighbours. Usually, periodic boundary conditions are taken in order to avoid spurious surface effects. During a Monte Carlo random walk, thermodynamic quantities (magnetization, susceptibility, etc.) are sampled and averaged over the number of steps. In this way it is possible, for instance, to locate the Curie temperature T_{C} of a ferromagnetic-paramagnetic phase transition by the drop of magnetization or by the susceptibility peak. Since the simulation supercell is finite, the magnetization does not fully disappear, and the susceptibility peak overestimates somewhat T_{C}. However, there are ways of correcting for this deficiency, by increasing the supercell size and using scaling arguments.[22] As an

Material	T_C(K) (exp)	T_C(K) (RPA)	T_C(K) (mean-field)	Ref.
Fe bcc	1044	950	1414	(a)
Co fcc	1403	1311	1645	(a)
Ni fcc	624	350	397	(a)
NiMnSb	730	900	1112	(b)
CoMnSb	490	671	815	(b)
Co_2CrAl	334	270	280	(b)
Co_2MnSi	985	740	857	(b)

Table 1. Experimental and calculated Curie temperatures (in Kelvin, within the RPA) of various ferromagnetic materials. Calculated values taken from: (a): Pajda et al,[25] (b): Sasioglu et al.[26]

example we show in fig. 2 the temperature-dependent magnetization and susceptibility of fcc Co calculated within the Monte Carlo method.

Mean-field theory is a physically transparent and computationally trivial way of estimating thermodynamic properties, however it lacks accuracy because it neglects fluctuations. As regards the Curie temperature, it is systematically overestimated by mean-field theory (assuming applicability of the Heisenberg model). Given the exchange interactions $J_{nn'}$ the mean-field result for T_C in a monoatomic crystal has the simple form

$$k_B\, T_C = \frac{1}{3} M^2 \sum_{n'} J_{nn'}. \tag{32}$$

Another widely used method for estimating the Curie temperature is the random phase approximation. It yields results much improved with respect to mean-field theory with only little increase of the computational burden. It is based on the Green function method for the quantum Heisenberg model, where a decoupling is introduced in the Green function equation of motion, as proposed by Tyablikov for $s = \frac{1}{2}$ systems.[23] Further refinements[24,25] of the RPA for higher-spin systems allow the transition to the classical limit by taking $s \to \infty$. The Curie temperature in a monoatomic lattice is then given by

$$\frac{1}{k_B\, T_C} = \frac{3}{2} \frac{1}{N} \sum_{\vec{q}} \frac{1}{E(\vec{q})} \tag{33}$$

where $E(\vec{q})$ is the magnon (or spin-spiral) energy, calculated by a Fourier transform of $J_{nn'}$ or directly by constrained DFT, and N the number of atoms in the system. For multisublattice systems, a modified version of RPA can be used.[26]

7.2 Time-dependent magnetic properties and Landau-Lifshitz spin dynamics

In case that one is interested in the time dependence of the magnetic moments under the influence, e.g., of an external field pulse, the method of magnetization dynamics can be used. The classical equations of motion associated with this method are the Landau-Lifshitz equations for the moments $\{\vec{M}_n\}$,

$$\frac{d\vec{M}_n}{dt} = \vec{H}_n^{\text{eff}} \times \vec{M}_n, \tag{34}$$

$$\vec{H}_n^{\text{eff}} = \sum_{n'} J_{nn'} \vec{M}_{n'} + \vec{H}^{\text{ext}}. \tag{35}$$

These are first-order equations in time which describe the precession of the magnetic moment due to external fields (different than an electric dipole, which will rotate towards the direction of an electric field, the magnetic dipole is essentially an angular momentum and therefore will precess around a magnetic field). The effective field defined in eq. (35) comprises the exchange interaction with the neighbours and an externally applied magnetic field. However, other terms can be included in $\vec{H}_n^{\rm eff}$, such as the magnetocrystalline anisotropy or the magnetic field created by the very moments of the material itself—the latter becomes most important in large ferromagnetic systems, and we discuss in the next subsection.

As is obvious by taking the dot product of eq. (34) with $\vec{M}_n$, $\vec{M}_n \cdot d\vec{M}_n/dt = 0$, i.e., the Landau-Lifshitz equations conserve the magnitude of the moments. They also conserve the total energy. However, dissipation effects that lead to damping of the precession can be taken into account by an additional phenomenological term of the form $\lambda(\vec{H}_n^{\rm eff} \times \vec{M}_n) \times \vec{M}_n$, where a parameter λ describes the damping strength. Temperature effects can also be simulated by additional phenomenological terms of stochastic forces, through an approach similar to Langevin molecular dynamics.[27]

We should note here the existence of a formalism for fully ab-initio spin dynamics, i.e., without the assumption of a Heisenberg model.[28] (From this formalism the Landau-Lifshitz equations follow as a limiting case.) However, this approach is computationally heavy, as it requires self-consistent density functional calculations at each time step of the system evolution.

7.3 Dipolar field calculation and related multiscale modelling

We now discuss the effect of the dipole-dipole interaction on the magnetic configuration. By this we mean the interaction of each magnetic dipole (here, atomic magnetic moment) with the magnetic field created by all other dipoles in the system. It is well-known that the this type of interaction between two moments $\vec{M}_n$ and $\vec{M}_{n'}$, connected by a vector $\vec{R}_{nn'} = \vec{R}_n - \vec{R}_{n'}$, has the form

$$E_{\rm dip}(\vec{R}) = \frac{3\,(\vec{M}_n \cdot \vec{R}_{nn'})(\vec{M}_{n'} \cdot \vec{R}_{nn'}) - (\vec{M}_n \cdot \vec{M}_{n'})\,R_{nn'}^2}{R_{nn'}^5} \tag{36}$$

Equivalently, each moment feels a magnetic field, the *dipolar field* $\vec{H}_n^{\rm dip}$, to be included in $\vec{H}_n^{\rm eff}$ in the Landau-Lifshitz equation, of the form

$$\vec{H}_n^{\rm dip} = \sum_{n' \neq n} \frac{3\,\vec{R}_{nn'}(\vec{M}_{n'} \cdot \vec{R}_{nn'}) - \vec{M}_{n'}\,R_{nn'}^2}{R_{nn'}^5} \tag{37}$$

Compared to the nearest-neighbour exchange interactions $J_{nn'}$, the interaction between two dipoles is weak, but the complication is that the summation (37) cannot be restricted to a few neighbours only, as it falls off relatively slowly with distance ($\sim 1/R_{nn'}^3$). Especially in three-dimensional systems the sum is guaranteed to converge only by finite-size effects of the sample, i.e., it becomes a meso- or macroscopic property and the sample boundaries become important.[a] $\vec{H}_n^{\rm dip}$ is evidently time-consuming to calculate; particularly a brute-force calculation would be impossible for large systems. There are, however,

[a] In large ferromagnetic systems the dipolar field cannot be neglected, as it is responsible for the emergence of magnetic domains.

special techniques that allow for a fast, approximate calculation of $\vec{H}_n^{\rm dip}$. This is even more crucial for spin dynamics, as $\vec{H}_n^{\rm dip}$ depends on the moments configuration and has to be calculated anew at each time step.

One such technique is the *fast multipole method*, originally introduced to treat the problem of Coulombic interactions.[29] The central idea is to divide space in regions of different sizes, and treat the collective field from each region by a multipole expansion up to a certain order. The higher the order, the more accurate and expensive the calculation. Given a certain expansion order, regions that are far away from the point of field-evaluation can be large, while regions that are close have to be smaller to maintain accuracy (the criterion of region size is the opening angle D/R, with D the diameter of the region and R its distance from the point of field-evaluation). An essential ingredient of the fast multipole method is the efficient derivation of multipoles of a large region from the multipoles of its subregions. This derivation requires the calculation of multipole expansion and translation coefficients, which, however, depend only on the geometry and for magnetic systems have to be evaluated only once (as the magnetic moments are not moving).

A fast evaluation of the dipolar field allows for multiscale simulations in magnetostatics[30] or magnetization dynamics, also in a sense that we have not discussed up to this point. In such simulations, the transition from the large (mesoscopic or even macroscopic) scale to the atomic scale is done in a seamless way. The idea is to treat the magnetization as a continuous field by a coarse grained approach in regions where it is relatively smooth, whereas to gradually refine the mesh, even up to the atomic limit, in regions where the spatial fluctuations become more violent (e.g. magnetic vortex cores, Bloch points, monoatomic surface step edges, ferromagnet-antiferromagnet interfaces, etc.). In the continuum limit, however, the Landau-Lifshitz equations (34) must be rewritten in terms of a continuous magnetization $\vec{M}(\vec{r})$ and the *spin stiffness* A:

$$\frac{d\vec{M}(\vec{r})}{dt} = \vec{H}^{\rm eff}(\vec{r}) \times \vec{M}(\vec{r}) \tag{38}$$

$$\vec{H}^{\rm eff}(\vec{r}) = \frac{2}{M_s^2} A \nabla^2 \vec{M}(\vec{r}). \tag{39}$$

$M_s = |\vec{M}(\vec{r})|$ is the absolute value of the continuum magnetization (also called saturation magnetization in ferromagnetic samples). The term $A\nabla^2\vec{M}(\vec{r})$ results from taking $\sum_{n'} J_{nn'}\vec{M}_{n'}$ to the continuum limit; the spin stiffness is given (in an example of a monoatomic crystal with atomic moment M and primitive cell volume V_c) in terms of the exchange constants as $A = (1/4V_c)M^2 \sum_n J_{0n} R_n^2$, with R_n the distance of atom n from the origin.

8 Back-Coupling to the Electronic Structure

So far we have discussed how the transition from the DFT to the Heisenberg model is achieved by fitting the Heisenberg model parameters to DFT total energies at and close to the ground state. However, at higher temperature (close to the Curie temperature, that can be of the order of 1000 K) the local electronic structure can change. Several mechanisms can contribute to this: lattice vibrations, single-electron excitations, collective electronic

excitations such as magnons, structural phase transitions (such as the hcp to fcc transition of Cobalt above 700 K) etc. As a consequence, the pair exchange parameters $J_{nn'}$ calculated from the low-temperature electronic structure could be significantly altered.

Perhaps the most serious effect can be caused by the non-collinear magnetic configurations at high temperature, in which the angle between first-neighbouring moments can be of the order of 30°. At such high angles, and depending on the system, the parametrization of the total energy with respect to the Heisenberg model can be insufficient—recall that, in principle, the Heisenberg hamiltonian is justified as the lowest-order term in an expansion with respect to the angle. An often encountered consequence of an altered local electronic structure is a change of the atomic moments. Furthermore, as this angle is not static, but fluctuating in time, it is no use to simply perform static non-collinear calculations at this angle and derive the $J_{nn'}$ by small deviations. We are thus faced with the problem of a back-coupling of the high-temperature state to the electronic structure; i.e., of approximating the local electronic properties in the presence of thermal fluctuations.

Two solutions are frequently used to this problem. The first is to go beyond the Heisenberg model and perform a more thorough parametrization of the energy as a function of the moments, including also possible changes in the magnitude of the moments. This method has been applied, e.g., by Uhl and Kübler.[31] The disadvantage is that it can be computationally expensive, both due to the number of self-consistent constrained-DFT calculations required for a parametrization of the multi-dimensional space $\{\vec{M}_n\}$, and because of the more involved Monte Carlo calculations where the change of the moments magnitude has to be accounted for. There are, however, reasonable approximations that can reduce the necessary number of parameters, while the Curie temperature can be found within a modified mean-field theory.[31]

The second solution is to assume that the Heisenberg model is still adequate to describe the phase transition, but with "renormalized" parameters, chosen such that the change of the local electronic structure is taken into account by an averaging over angles. This solution is intuitive but certainly not rigorous. It is, however, simple to include within Green function electronic structure methods, by assuming an analogy of the high-temperature state with a spin-glass state and employing the coherent potential approximation (CPA). Spin-glass systems are characterized by *disordered local moment* (DLM) states, consisting of two different magnetic "species" that correspond, say, to the magnetic moment pointing "up" (A) or "down" (B). These are encountered in a random manner with a probability $1-x$ and x, respectively: the DLM state is of the form $A_{1-x}B_x$. For $x=0$ we recover the ferromagnetic state, while for $x=0.5$ we have complete magnetic disorder. (Note that a DLM state is different than the antiferromagnetic state, in which the species A and B are well-ordered in two sublattices.) Under the assumption of an analogy of high-temperature states in ferromagnets to DLM systems, the ferromagnet at the Curie temperature is approximated by the alloy $A_{0.5}B_{0.5}$.

The CPA is a method for the description of chemical disorder in alloys, and can be applied here to the magnetic type of alloy $A_{0.5}B_{0.5}$. Within the CPA, the Green function $\bar{\boldsymbol{G}}$ and scattering matrix $\bar{\boldsymbol{t}}$ of an effective average medium are sought, such that the additional scattering of atoms A and B in this medium vanishes on the average. We skip the derivation, which can be found in many textbooks,[32,33] and give only the final CPA condition

that has to be fulfilled:

$$\bar{\boldsymbol{t}}^{-1} = (1-x)\,\boldsymbol{t}_A^{-1} + x\,\boldsymbol{t}_B^{-1} + (\bar{\boldsymbol{t}}^{-1} - \boldsymbol{t}_A^{-1})(\bar{\boldsymbol{t}}^{-1} - \bar{\boldsymbol{G}})^{-1}(\bar{\boldsymbol{t}}^{-1} - \boldsymbol{t}_B^{-1}), \tag{40}$$

$$\bar{\boldsymbol{G}} = \boldsymbol{g}\,\left(1 - \bar{\boldsymbol{t}}\boldsymbol{g}\right)^{-1} \tag{41}$$

with $\boldsymbol{g}$ the free-space structural Green function in the KKR formalism.[16] Expression (41) is the Dyson equation for the Green function of the average medium, which depends on the average-medium scattering matrix $\bar{\boldsymbol{t}}$. The latter is determined by Eq. (40), which contains $\bar{\boldsymbol{t}}$ also on the right-hand side (explicitly and also through $\bar{\boldsymbol{G}}$), and is solved self-consistently by iteration. At the end, the Green functions of species A and B are projected out from the average medium Green function again via the Dyson equation

$$\boldsymbol{G}_{A,B} = \bar{\boldsymbol{G}}\,\left(1 - (\boldsymbol{t}_{A,B} - \bar{\boldsymbol{t}})\,\bar{\boldsymbol{G}}\right)^{-1} \tag{42}$$

and used for the calculation of the electronic structure of the two atomic species.

Given the CPA Green function for the $A_{0.5}B_{0.5}$ DLM state, the method of infinitesimal rotations can be employed to obtain the pair exchange constants. Assuming that the DLM state represents the magnetic structure at the Curie temperature, the exchange constants obtained by this method should be more appropriate to use in the Heisenberg hamiltonian close to T_{C} than the ones obtained from the ground state. However, this is not guaranteed, especially in view of the fact that the CPA neglects the short-range magnetic order that is present even at T_{C}.

9 Concluding Remarks

The Multiscale Programme discussed here is widely used today, however, the matter is surely not closed. Mainly two types of difficulties are present and are the subject of current research. First, density functional theory within the local spin density or generalized gradient approximation is not able to describe the ground state properties of every material. When electron correlations (on-site electron-electron repulsion and temporal electron density fluctuations) become particularly strong, these approximations fail. Characteristic of such problems are f-electron systems, transition metal oxides or molecular magnets. Improved concepts exist and are applied, such as the LSDA+U or dynamical mean-field theory, however, at the moment these methods rely on parameters that cannot always be found without a fit to experiment.

Second, the excited state properties are also not always accessible to density functional theory. The adiabatic hypothesis, together with constrained DFT, work up to a point, but effects as the magnon lifetime or frequency-dependent interactions are neglected. Current work in this direction is done within approximations as the GW or time-dependent DFT, with promising results. These methods are, however, still computationally very expensive, and the extent of improvement that they can offer to the calculation of thermodynamical properties remains unexplored.

Acknowledgments

I am grateful to D. Bauer, S. Blügel, P.H. Dederichs, R. Hertel, M. Ležaić, S. Lounis, and L.M. Sandratskii for illuminating discussions on the subjects presented in this work.

References

1. P. Hohenberg and W. Kohn Phys. Rev. **136**, B864 (1964).
2. W. Kohn and L. J. Sham Phys. Rev. **140**, A1133 (1965).
3. U. von Barth and L. Hedin, J. Phys. C: Solid State Phys. **5**, 1629 (1972).
4. V.L. Moruzzi, J.F. Janak, and A.R. Williams, *Calculated electronic properties of metals* (Pergamon, New York 1978).
5. D.M. Ceperley and B.J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
6. S.H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
7. See, for example, J.P. Perdew, K. Burke, and Y. Wang, Phys. Rev. B **54**, 16533 (1996); J.P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
8. V.I. Anisimov, F. Aryasetiawan, and A.I. Lichtenstein J. Phys.: Condens. Matter **9**, 767 (1997).
9. P.H. Dederichs, R. Zeller, H. Akai, and H. Ebert, J. Magn. Magn. Matter **100**, 241 (1991).
10. P.H. Dederichs, S. Blügel, R. Zeller, and H. Akai, Phys. Rev. Lett. **53**, 002512 (1984).
11. A. Oswald, R. Zeller, and P. H. Dederichs, J. Magn. Magn. Mater. **54-57**, 1247 (1986); N. Stefanou and N. Papanikolaou, J. Phys.: Condens. Matter **5**, 5663 (1993).
12. A.R. Mackintosh and O.K. Andersen, *The electronic structure of transition metals*, in: *Electrons at the Fermi surface*, Ed. by M. Springfold (Cambridge University Press, Cambridge, 1980); M. Weinert, R. E. Watson, and J. W. Davenport, Phys. Rev. B **32**, 2115 (1985); V. Heine, Solid State Phys. 35, 1 (1980); M. Methfessel and J. Kübler, J. Phys. F 12, 141 (1982).
13. A. Oswald, R. Zeller, P.J. Braspenning, and P.H. Dederichs, J. Phys. F **15**, 193 (1985).
14. L.M. Sandratskii, Phys. Stat. Sol. (b) 135, 167 (1986).
15. S. V. Halilov, H. Eschrig, A. Y. Perlov, and P. M. Oppeneer, Phys. Rev. B **58**, 293 (1998).
16. P. Mavropoulos and N. Papanikolaou, *The Korringa-Kohn-Rostoker (KKR) Green Function Method I. Electronic Structure of Periodic Systems*, in *Computational Nanoscience: Do It Yourself!*, eds. J. Grotendorst, S. Blügel, and D. Marx, Winterschule, 14.-22. Februar 2006, (Forschungszentrum Jülich 2006); `http://www.fz-juelich.de/nic-series/volume31/volume31.html`
17. P.H. Dederichs, S. Lounis, and R. Zeller, *The Korringa-Kohn-Rostoker (KKR) Green Function Method II. Impurities and Clusters in the Bulk and on Surfaces*, in *Computational Nanoscience: Do It Yourself!*, eds. J. Grotendorst, S. Blügel, and D. Marx, Winterschule, 14.-22. Februar 2006, (Forschungszentrum Jülich 2006); `http://www.fz-juelich.de/nic-series/volume31/volume31.html`
18. S. Lounis, Ph. Mavropoulos, P. H. Dederichs, and S. Blügel Phys. Rev. B **72**, 224437 (2005).
19. A.I. Liechtenstein, M.I. Katsnelson, V.P. Antropov, and V.A. Gubanov, J. Magn. Magn. Mater. **67** 65 (1987).
20. P. Bruno, Phys. Rev. Lett. **90**, 087205 (2003).
21. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
22. D.P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press (2000).

23. S.V. Tyablikov, *Methods of Quantum Theory of Magnetism* (Plenum Press, New York, 1967).
24. H. B. Callen, Phys. Rev. **130**, 890 (1963).
25. M. Pajda, J. Kudrnovsky, I. Turek, V. Drchal, and P. Bruno, Phys. Rev. B **64**, 174402 (2001).
26. E. Sasioglu, L. M. Sandratskii, P. Bruno, and I. Galanakis, Phys. Rev. B **72**, 184415 (2005).
27. See, for example, V.P. Antropov, S.V. Tretyakov, and B.N. Harmon, J. Appl. Phys. **81**, 3961 (1997).
28. V. P. Antropov, M. I. Katsnelson, B. N. Harmon, M. van Schilfgaarde, and D. Kusnezov, Phys. Rev. B **54**, 1019 (1996).
29. L. Greengard and V. Rokhlin, J. Comput. Phys. 73, 325 (1987).
30. Thomas Jourdan, Alain Marty, and Frederic Lancon, Phys. Rev. B **77**, 224428 (2008). A. Desimone, R.V. Kohn, S. Müller, and F. Otto, Multiscale Modeling and Simulation 1, **57** (2003).
31. M. Uhl and J. Kübler, Phys. Rev. Lett. **77**, 334 (1996).
32. J. Kübler, *Theory of Itinerant Electron Magnetism*, Oxford University Press, 2000.
33. A. Gonis, *Theoretical Materials Science*, Materials Science Society, 2000.

First-Principles Based Multiscale Modelling of Alloys

Stefan Müller[1,2]

[1] Lehrstuhl für Theoretische Physik 2
Universität Erlangen-Nürnberg
Staudtstrasse 7, 91058 Erlangen, Germany

[2] Lehrstuhl für Festkörperphysik
Universität Erlangen-Nürnberg
Staudtstrasse 7, 91058 Erlangen, Germany
E-mail: stefan.mueller@physik.uni-erlangen.de

Although modern computer codes based on density functional theory (DFT) allow the reliable prediction of many surface and bulk properties of solids, they cannot be applied, when the problem of interest demands a consideration of huge configuration spaces or model systems containing many thousand atoms. Important examples are precipitation and segregation in metal alloys where substitutional ordering phenomena on a mesoscopic scale are involved. Moreover, in general first-principles methods based on DFT do not allow for exchange processes between atoms and therefore, do not consider configurational enthalpies being a prerequisite for modelling the temperature-dependence of decomposition reactions or segregation phenomena. In this contribution, recent developments, possibilities and limitations to study ordering phenomena and ground-state properties based on first-principles methods will be discussed. It will be demonstrated how the combination of DFT calculations with so-called Cluster Expansions and Monte-Carlo simulations allows for a quantitative prediction of alloy properties from the miscroscopic up to the meso-, and even macroscale without any empirical parameters.

1 Introduction: The Definition of "Order"

If A- and B-atoms are forced to crystallize on a common lattice, they may either order (AB-bonds) or cluster (AA- and BB-bonds) depending on whether the occupation of neighboring lattice sites by identical or different species is energetically favoured. However, the situation becomes more complex, when temperature comes into play: The temperature-composition phase diagram of a binary solid state alloy, $\mathrm{A}_{1-x}\mathrm{B}_x$, may consist of homogeneous single-phase regions (such as ordered compounds A_mB_n) as well as heterogeneous, phase-coexistence regions[1]. Besides intermetallic compounds, i.e. long-range ordered (LRO) phases, which are mostly observed at low temperatures, in many binary metal systems so-called "solid solutions" exist. Although such solid solutions are often described by a lattice grid randomly occupied by A and B atoms, more or less *all* solid solutions show substitutional short-range order (SRO). Indeed, SRO may have a tremendous influence on the energy and stability of this alloy phase. Consequently, the physical properties of solid solutions must be modelled by a *disordered* alloy which is not necessarily a *random* alloy. In fact, SRO makes a quantitative, theoretical description of alloys on a quantum-mechanical basis rather complex. One may ask, if it is really necessary to consider SRO for a quantitative description of an alloy's stability. Fo this, we consider the solid solution of α-brass (Cu-rich Cu-Zn). For this phase, it well-known from experiment[2] and theory[3] that characteristic SRO occurs. This can be seen in Figure (1) which compares calculated mixing enthalpies[3], $\Delta H_{mix}(x,T)$, for different temperatures with experimental data

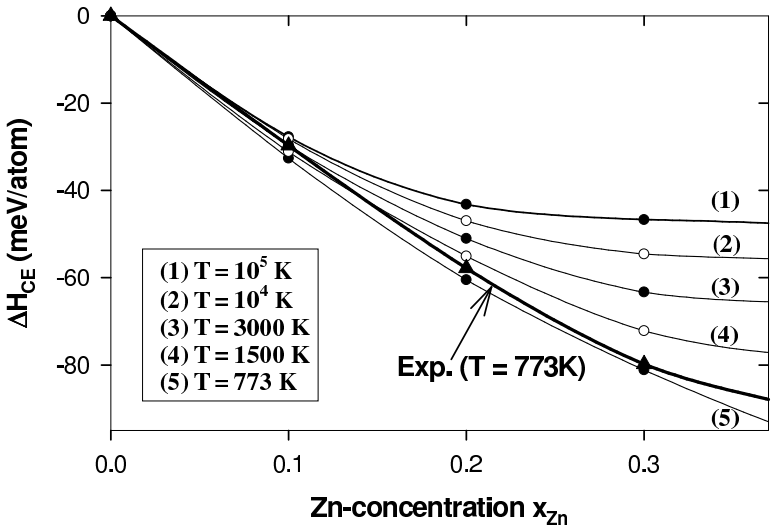


Figure 1. Calculated mixing enthalpies of α-brass for different temperatures[3], in comparison with experimental data[4] (bold line).

taken from Hultgren's book[4]. The mixing enthalpy, ΔH_f^{DFT} per atom of configuration σ is defined as

$$\Delta H_{mix}(\sigma) = \frac{1}{N} E^{tot}(A_{1-x}B_x, \sigma) - x E_A^{tot}(a_A) - (1-x) E_B^{tot}(a_B) \qquad (1)$$

with N being the total number of atoms in the disordered alloy. $E^{tot}(A_{1-x}B_x, \sigma)$ is the total energy of the geometrically fully relaxed configuration σ with concentration x of B-atoms ($0 \leq x \leq 1$). Furthermore, a_A and a_B are the equilibrium lattice constants of the elements A and B, $E_A^{tot}(a_A)$ and $E_B^{tot}(a_B)$ are the respective total energies. Since all total energy values are negative, a positive sign of ΔH_{mix} stands for phase-separation, while a negative sign of ΔH_{mix} stands for ordering (as in the case of α-brass). The theoretical calculations in Figure (1) are performed by combining density functional theory with methods from statistical physics which will be explained in the next section: We start with the random alloy ($T \to \infty$) and go down to temperatures where short-range order sets in. Figure 1 shows that the calculation neglecting ordering phenomena ($T = 10^5$K, corresponding to a random alloy) leads to much higher mixing enthalpies than in experiment. For higher Zn concentrations a good agreement between experiment and calculated mixing enthalpies can only be reached, if ordering phenomena are taken into account.

Before we discuss, how to calculate SRO, we need a a measure *how* to quantify it. Ziman[5] nicely described the difficulty to handle ordered zones in a disordered matrix by Fig. (2): For the given configuration, we cannot decide, if the atom marked by an arrow belongs to a "cluster of pure A-atoms" or to a "region of perfect AB-order". He demonstrated by applying percolation theory that almost every A-atom belongs to an infinite cluster of A atoms. Paradoxically, if we are looking for ordered domains (Fig. (2)), then almost every atom belongs to an infinite domain with perfect AB-ordering. Help comes by introducing statistical concepts[5–7]: For a system consisting of N sites each surrounded by M neigh-

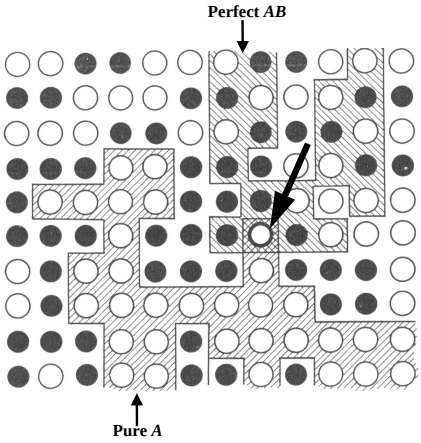


Figure 2. The dilemma in describing ordering (taken from Ziman[5]): Does the atom marked by an arrow belong to "a cluster of pure A-atoms", or to a "region of perfect AB-order"?

bors, the probability of a bond being of AB-type is given by

$$P_{AB} = lim_{N\to\infty}\left(\frac{N_{AB}}{\frac{1}{2}MN}\right) \tag{2}$$

with N_{AB} being the total number of AB-type bonds. The denominator gives the total number of bonds in the system. If we assume that each site of the system is independently occupied by an A- or B-atom with probability x_A or x_B ($x_A + x_B = 1$), then P_{AB} would be $2x_Ax_B$. Then, the nearest-neighbor correlation parameter Γ_{AB} can be defined as

$$\Gamma_{AB} = \frac{1}{2}P_{AB} - x_Ax_B.$$

Dividing Γ_{AB} by $-x_Ax_B$ leads to the well-known *Warren-Cowley short-range order parameter*[8]

$$\alpha_j = 1 - \frac{P^j_{AB}}{2x_Ax_B}. \tag{3}$$

Here, α_j is already extended to arbitrary neighbor distances j. The sign of α_j indicates whether atoms in a given distance j prefer AB-ordering ($\alpha_j < 0$) or clustering ($\alpha_j > 0$). The SRO parameter are normalized such that $-1 \leq \alpha_j \leq +1$. Since α_j can be determined from diffuse X-ray and neutron diffraction experiments[9–11], a *quantitative* comparison between calculation and measurement is possible.

If we wish to describe and understand the properties of different solid phases and their stability on a quantum-mechanical basis, we have to solve three fundamental problems (Fig. 3):

(i) *The configurations-space problem:* In general, first-principles calculations only considers atomic relaxations in the unit cell, but do not allow for exchange processes between

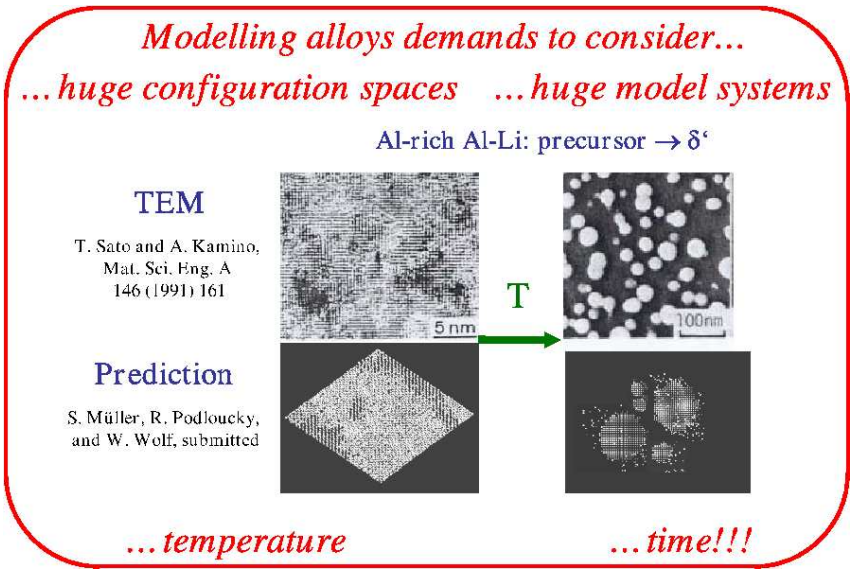


Figure 3. Comparison between predicted and measured precipitation in Al-rich Al-Li alloys. The theoretical description demands to overcome the four fundamental problems in materials modelling.

individual atoms. The latter is a prerequisite for an efficient and reliable ground-state search, i.e. for finding the configuration being lowest in energy for a given concentration.
(ii) *The multiscale problem:* The quantitative prediction of short-range order phenomena often requires models with giant unit cells. Model systems containing up to 10^6 atoms may be demanded, i.e. much more than the about 500 metal atoms treatable by today's computers.
(iii) *The temperature problem:* The temperature-dependence of ordering phenomena must not be neglected. However, in general, electronic structure theories are constructed to study $T = 0$K properties.

In principle, there is a fourth problem, namely the fact that many properties of alloys are not understandable in the framework of thermodynamics. In order to go beyond equilibrium properties of the system, kinetic approaches have to be considered. As a consequence the system's properties become *time-dependent*. As will be demonstrated in section 2, it is not an easy task to transform kinetic simulation results into a real-time scale.

The main aim of this lecture is to study the bulk and surface properties of metal alloys *without* any experimental parameters as input. As already mentioned in the Introduction, we use Density Functional Theory (DFT)[12,13] as starting point for our studies. Although DFT permits one to calculate alloy properties with an accuracy that often allows for a quantitative comparison with experimental data, it is usually limited to a small subset of the configuration space. The geometric relaxation of unit cells consisting of more than 100 atoms already becomes extremely difficult, and even impossible for some cases. So, compared to the 2^N configurations of a binary system containing N atoms, we are restricted to a very small part of the parameter space. Normally, a set of "intuitive structures" is chosen and that with the minimal energy is postulated as ground-state configuration. This,

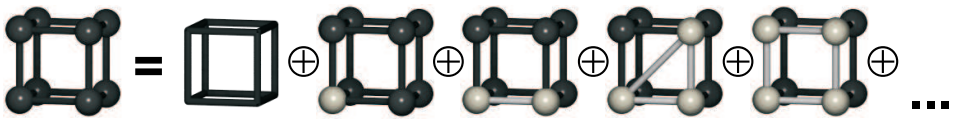


Figure 4. The concept of cluster expansions: The crystal is separated in characteristic figures (here, shown for the fcc-lattice). The energy of any configuration can then be written as linear combination of the characteristic energies J_f of the figures.

however, fails to allow for surprises, only one of the chosen input structures can result as ground-state. In order to circumvent this problem, the accuracy of DFT is extended to huge configuration spaces by combining DFT with concepts from statistical mechanics. The basic idea by Sanchez, Ducastelle and Gratias[14] is called "Cluster Expansion" (CE), and sketched in Fig. (4): For a given underlying lattice, the crystal structure is divided into characteristic figures such as pairs, triangles, etc. Then, the energy of *any* configuration σ on this lattice can be uniquely written[14] as linear combination of the characteristic energies J of each individual figure. In practice, the only error we make is that the sum must be truncated at some point. The Π_f's in Fig. (4) are structure-dependent factors and will be discussed in detail in section 2.2.

2 Methods

2.1 Elastic properties of alloys from density functional theory

Density functional theory (DFT) represents the probably most important many-particle approach in solid-state physics with respect to applications. Since there exists a number of excellent review articles (see e.g.[15,16]) and books (see e.g.[17–19]) about DFT, only some general remarks will be given: DFT is based on the Hohenberg-Kohn-theorem[12] stating that the energy of a system of interacting electrons in an external potential depends only on the ground state electronic density. In our case, namely the investigation of solid structures, the external potential is the Coulomb potential caused by the nuclei in a solid. The ground-state density can in principle be calculated from a variation ansatz, i.e. without any Schrödinger-equation, however for treating real problems the variational approach is unpracticable. Help came in 1965 by Kohn and Sham[13] who showed that the density wanted is given by the self-consistent solution of a set of single particle equations, called Kohn-Sham equations:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{e-nuc}(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r})\right]\Psi_i(\mathbf{r}) = \epsilon_i\Psi_i(\mathbf{r}) \qquad (4)$$

In this Schrödinger-like equation, the first term on the left side represents the kinetic energy operator, V_{e-nuc} the Coulomb potential due to the nuclei, V_H the Hartree potential, and V_{XC} is the exchange correlation potential. The latter comes from replacing the kinetic energy of interacting particles by that of non-interacting particles (which can be treated exactly) plus a term containing all correlation and exchange effects (which is unknown, but small compared to the other energy contributions). Well-known approximations for

V_{XC} are the Local Density Approximation (LDA)[20,21] and the Generalized Gradient Approximation (GGA)[22] . In LDA, the energy density of the inhomogeneous system is approximated by the density of the *homogeneous* electron gas which possesses exactly the same density as the actual *inhomogeneous* system. Although this sounds like a very rough approximation, especially for systems with strongly varying density, it works astonishing well for a huge number of problems. In GGA, additionally the gradient of the density is considered which can be important for systems where $n(\mathbf{r})$ changes dramatically with $\mathbf{r}$.

In practice, we can distinguish between more or less two different types of strategies: Methods using complex, but efficient basis sets for the wavefunctions, as the Linearized Augmented Planewave method (LAPW) and methods based on so-called pseudopotentials (PP) using plane waves as basis set (for a survey see e.g. the book by Singh[23]). The concept of pseudopotentials is roughly spoken that most physical properties of a solid are determined by the valence electron structure. Then, the number of plane waves necessary to describe the system can be tremendously decreased by replacing core electrons and ionic potential by a pseudopotential which is energetically much weaker and corresponds to a node-free wavefunction. Thereby, the pseudopotential has to fulfil the conditions that (a) the scattering properties of the elements are conserved and (b) outside the core-region pseudopotential and pseudo-wavefunction are identical to the corresponding full potential and wavefunction. Until some years ago, it was a very delicate task to study transition-metals by "classical", norm-conserving pseudopotentials[24,25]. With the development of ultrasoft pseudopotentials[26,27] and more recently, so-called PAW-potentials ("Projector Augmented Wave")[28,29] concepts from LAPW entered in PP-codes and allow for an accurate and fast treatment of practically all metal-system by a plane wave basis set.

In many cases, results retrieved from DFT calculations are used as input for other numerical and analytic models to describe a certain class of properties of an alloy system. One important example is the use of the DFT energetics in elasticity theory in order to calculate the strain behaviour of metal alloys. In Subsection (2.2), we will see, how the following concept permits one to understand the size versus shape relation of characteristic microstructures in metal alloys. Since strain is determined by the mechanical behaviour of the system, we separate the two components by creating an interface in a well-defined orientation between A- and B-atoms and demand that the whole system act as a pseudomorphic, epitaxial system, i.e. there are no dislocations at the interface. The idea to compare a binary alloy with an epitaxial film/substrate system allows to specify two types of quantities[30]:
(i) The *hydrostatic deformation energy* $\Delta E_A^{hydro}(a)$ being the energy required to hydrostatically deform the solid element A to the lattice constant a of the alloy.
(ii) The *epitaxial strain energy* $\Delta E_A^{epi}(a, \hat{G})$, representing the energy of the elemental solid A epitaxially (or, biaxially) deformed to the "substrate" lattice constant a in the two directions orthogonal to $\hat{G}$ and *relaxed* along $\hat{G}$.
The ratio of these two energies defines the *epitaxial softening function*[30,31]

$$q(a, \hat{G}) = \frac{\Delta E_A^{epi}(a, \hat{G})}{\Delta E_A^{hydro}(a)}. \tag{5}$$

Since it is always easier to deform a material epitaxially (biaxially) than hydrostatically (triaxially), $q \leq 1$. Small values of $q(a, \hat{G})$ indicate elastically soft directions $\hat{G}$. As an example, Fig. (5)(b) shows calculated softening functions, $q(a, \hat{G})$, for the fcc-elements Al

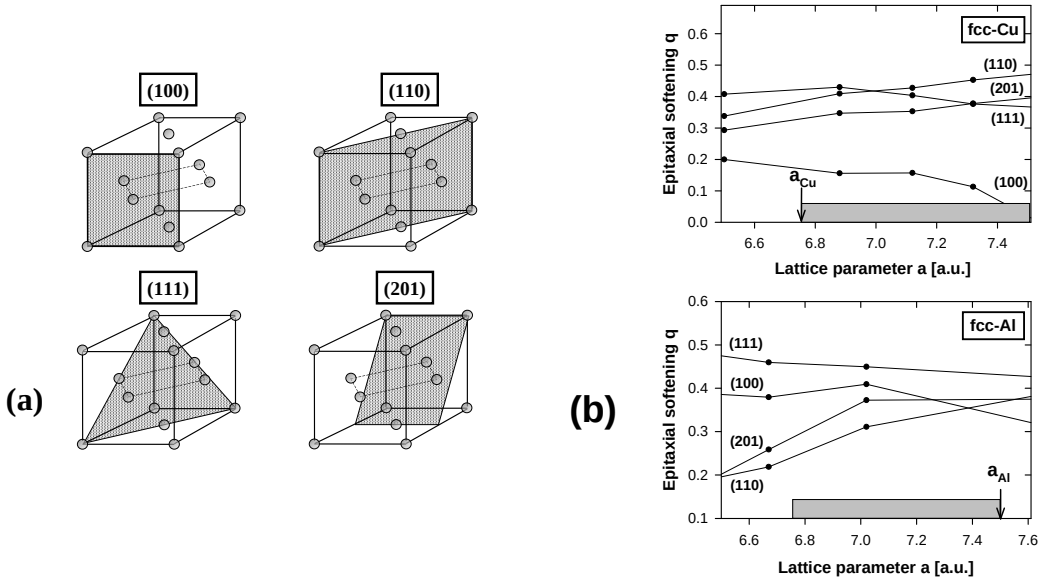


Figure 5. (a) Low index crystal orientations of the fcc-lattice indicated by hatched areas. (b) Epitaxial softening function $q(a, \hat{G})$, Eq. (5), for Cu and Al calculated via LDA. The shaded areas mark the lattice parameter range between the two components of the corresponding alloy. Arrows denote the position of the equilibrium lattice constant a_{eq} of each element. The lines are drawn merely to guide the eye.

and Cu along the crystal directions indicated in Fig. (5)(a). Obviously, the crystallographic order of elastic softness can change as function of the lattice parameter. For example, an only 2% compression of Al (Fig. (5)(b)) is softer along (110) than along (100), while at the equilibrium the opposite is true. This clearly indicates that for a description of strain effects in metals, not only the direction dependence of strain (*anisotropic* strain effects), but also the dependence of strain on the lattice parameter (*anharmonic* strain effects) must be taken into account[32,33]. In the harmonic elasticity theory, q depends only on the direction $\hat{G}$, but *not* on the substrate lattice constant a[30,34,35]:

$$q_{harm}(\hat{G}) = 1 - \frac{B}{C_{11} + \Delta\gamma_{harm}(\hat{G})} \tag{6}$$

with bulk modulus $B = \frac{1}{3}(C_{11}+2C_{12})$ and anisotropy parameter $\Delta = C_{44}-\frac{1}{2}(C_{11}-C_{12})$. The harmonic constants C_{11}, C_{12}, and C_{44} can be easily calculated from first-principles calculations[30] and consequently, Δ and B, too. γ_{harm} is a geometric function of the spherical angles Θ (polar angle) and Φ (azimuth angle) formed by $\hat{G}$:

$$\begin{aligned}\gamma_{harm}(\Phi,\Theta) &= sin^2(2\Theta) + sin^4(\Theta)sin^2(2\Phi) \\ &= \frac{4}{3}\sqrt{4\pi}\left[K_0(\Phi,\Theta) - \frac{2}{\sqrt{21}}K_4(\Phi,\Theta)\right]\end{aligned} \tag{7}$$

Here, K_l is the Cubic harmonic of angular moment l. If anharmonic effects become im-

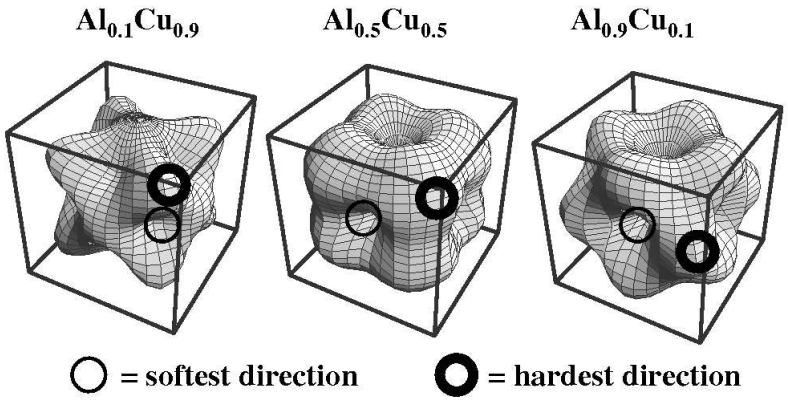


Figure 6. Parametric three dimensional presentation of the constituent strain ΔE_{CS}^{eq}, Eq. (10), of Al-Cu for compositions of 10%, 50%, and 90% Al. The distance from the surface to the centre of the cube represents the amount of the strain energy.

portant as in metal alloys, q additionally depends on the lattice parameter a:

$$\gamma(a,\hat{G}) = \gamma_{harm}(\hat{G}) + \sum_{l=0}^{l_{max}} b_l(a) K_l(\hat{G}). \tag{8}$$

This equation now also includes higher order cubic harmonics as necessary to go beyond the harmonic approximation (more details are given by Ozolins et al.[32]). Then, Eq. (6) becomes

$$q(a,\hat{G}) = 1 - \frac{B}{C_{11} + \Delta\gamma(a,\hat{G})}. \tag{9}$$

With $q(a,\hat{G})$ resulting from DFT calculations as displayed in Fig. (5)(b), the quantity $\gamma(a,\hat{G})$ can be taken from Eq. (9) and, in turn, the coefficients $b_l(a)$ results via Eq. (8). The determination of $b_l(a)$ permits one to generalize calculated epitaxial energies, $\Delta E_A^{epi}(a,\hat{G})$ for a discrete set of directions to *arbitrary* directions $\hat{G}$.

We will apply it to parameterize the equilibrium *constituent* (or *coherency*) *strain energy* $\Delta E_{CS}^{eq}(x,\hat{G})$ which is defined as the strain energy required to maintain coherency between a "piece" of material A and a "piece" of material B along an interface with orientation $\hat{G}$. This structure represents a so-called *superlattice* A_nB_n along a certain direction $\hat{G}$ with $n \to \infty$. In practice, the calculated elemental epitaxial energies are used to determine the constituent strain energy that is determined by the equilibrium value of the composition-weighted sum of the epitaxial energies of A and B:

$$\Delta E_{CS}^{eq}(x,\hat{G}) = \min_{a_p}[x\Delta E_A^{epi}(a_p,\hat{G}) + (1-x)\Delta E_B^{epi}(a_p,\hat{G})] \tag{10}$$

where $a_p(x)$ is the lattice constant that minimizes ΔE_{CS}^{eq} at each x. The constituent strain can be illustrated by a three-dimensional parametrization in terms of a sum of Kubic harmonics, as shown in Fig (6) for for three different Al-concentrations of the system Al-Cu.

Here, the distance from the surface to the centre of the cube represents the strain energy in this crystallographic direction. For $Al_{0.1}Cu_{0.9}$, we see that this distance is maximal along the body diagonal (marked by a bold circle), i.e. the crystallographic [111] direction, whilst the distance is shortest along the square face diagonal (marked by a thin circle), i.e. the [110] direction. With increasing Al composition the situation changes: $Al_{0.5}Cu_{0.5}$ owns the smallest constituent strain for [100], while [111] is still the hardest direction. For 90% Al, the figure has a "depression" in the very soft [100] direction, but a protrusion in the hard [111] direction. As we will see next, the concept of constituent strain is very important to describe morphological properties of alloys.

2.2 Controlling configuration space and length scales: The UNCLE code

As discussed in Section II, the idea of cluster expansions[14] is to express the atomically relaxed energy, $E(\sigma)$, of arbitrary lattice configurations σ on a given, underlying lattice as linear sum of energies characteristic of geometric figures, such as biatoms, triatoms, etc. (see Fig. (4)). To realize this idea, we transform the "alloy problem" to an Ising model. Each atom i of an $A_{1-x}B_x$ alloy is assigned to a spin-value $S_i = -1$, if i is an A-atom, and to $S_i = +1$, if i is a B-atom. Then, the energy of each configuration can be expressed by an Ising-expansion:

$$E(\sigma) = J_0 + \sum_i J_i S_i(\sigma) + \sum_{j<i} J_{ij} S_i(\sigma) S_j(\sigma) + \sum_{k<j<i} J_{ijk} S_i(\sigma) S_j(\sigma) S_k(\sigma) + ... \quad (11)$$

The first two terms on the right define the energy of the random alloy (with zero mutual interactions), the third term contains all pair interactions, the fourth all three-body interactions, etc. This equation can be brought to a compact form by introducing a correlation function $\bar{\Pi}_F$ for each class of symmetry-equivalent figures F[36]:

$$\bar{\Pi}_F(\sigma) = \frac{1}{N D_F} \sum_f S_{i_1}(\sigma) S_{i_2}(\sigma) ... S_{i_m}(\sigma) \quad (12)$$

Here, D_F gives the number of figures of class F per site. The index f runs over the ND_F figures in class F and m denotes the number of sites of figure f. Then, Eq. (11) becomes[35]

$$E(\sigma) = \sum_F D_F \bar{\Pi}_F(\sigma) J_F \quad (13)$$

The coefficients J_F of the cluster expansion are determined by fitting to an input database. This input database consists of a set of atomic configurations, whose energy has been determined, e.g., using ab-initio methods. An efficient cluster expansion method will facilitate the exchange of structural information between the fitting routines and the first-principles code. This decreases the amount of user time required and reduces the chances for human error.

Our new computer code UNCLE (UNiversal CLuster Expansion)[37] has been designed to adapt the output of the pseudopotential code VASP[27,28,38–40] and the FLAPW code FLAIR[41–43]. It should be mentioned though, that the source of the input values in the database can be arbitrary, and do not necessarily have to originate from first-principles calculations. For every input value in the database, the corresponding structural information is given as follows: real-space coordinates of the supercell B, the number of each

chemical atomic species in the cell, and positions of the basis atoms within the superstructure. The latter is given either in direct or Cartesian coordinates. Following the structural information, the corresponding value of the observable to be expanded is given.

After the input structures have been read in, UNCLE checks whether all their basis atoms lie on the lattice and whether there are symmetry-equivalent structures within the input list. As trivial as this step may seem, in practice this becomes an extremely useful feature; converged cluster expansions typically require around 50–150 input structures, which tend to contain subsets of similar, though symmetrically-distinct, atomic configurations. This can cause unintentional duplication of input structures, which not only wastes calculation time, but also falsely overweights the structure during the fitting.

The choice of atomic configurations, from which the effective cluster interactions are extracted, affects the ECIs. To avoid biasing the input database, and thus the ECIs, we systematically increase the database. We begin with a hand-chosen set $\{\sigma\}$ of usual suspects, small-unit-cell structures derived from the parent lattice, and some quasi-random structures. The first cluster expansion determined from this initial set makes predictions, perhaps not accurately, for the ground states and other structures with a "low" enthalpy of formation. One efficient tool to find structures with important "structure information" for the determination of the interactions is a ground-state search[44,45] in the early stage of the construction: For a "starting set" of about 20 DFT energies of arbitrary input-structures, a CE fit is performed. The resulting interactions are then used to predict the energy of *all possible structures* with e.g. up to 20 atoms per unit cell (the latter is indeed a very reasonable restriction, since most known stable structures in binary metal alloys own clearly less than 20 atoms per unit cell). Such an analysis based on Eq. (13) takes only some hours on a high-performance PC. Afterwards, the CE energies of all structures are plotted as function of composition, and a ground-state line is constructed. This is schematically shown in Fig. (7): An individual structure σ only contributes to the ground-state line, if the linear energy average between the stable structures at next higher and lower concentration is energetically less favourable than the energy of σ. More precisely, for three structures α, σ and β with $x(\alpha) < x(\sigma) < x(\beta)$ which are the lowest in energy for their individual concentrations, the structure σ has to fulfil the condition

$$E(\sigma) < \frac{x(\sigma) - x(\beta)}{x(\alpha) - x(\beta)} E(\alpha) + \frac{x(\sigma) - x(\alpha)}{x(\beta) - x(\alpha)} E(\beta) \tag{14}$$

in order to be the ground-state at $x(\sigma)$. If Eq. (14) holds, a mixture of the phases α and β would be higher in energy than structure σ. With the ground state line constructed, UNCLE automatically checks for all structures which lie on it, whether they are already considered as input structures for the CE. If not, their DFT energy is calculated and added to the input-structure set. This cycle is repeated, as shown in Fig. 8, letting the current cluster expansion itself pick new structures to add to the database.

In practice, the prediction of the energy (or any other observable) over a system's configuration space (e.g., ground state searches) by the help of UNCLE requires only minimal user input. We have implemented an algorithm[46] that automatically generates all possible atomic configurations within all geometrically possible supercells for an arbitrary number of basis atoms on a given lattice. The algorithm removes all symmetry-equivalent structures and still scales linearly with the number of unique configurations. For a ground state search based on the cluster-expansion Hamiltonian, Eq. (13), the user only has to provide

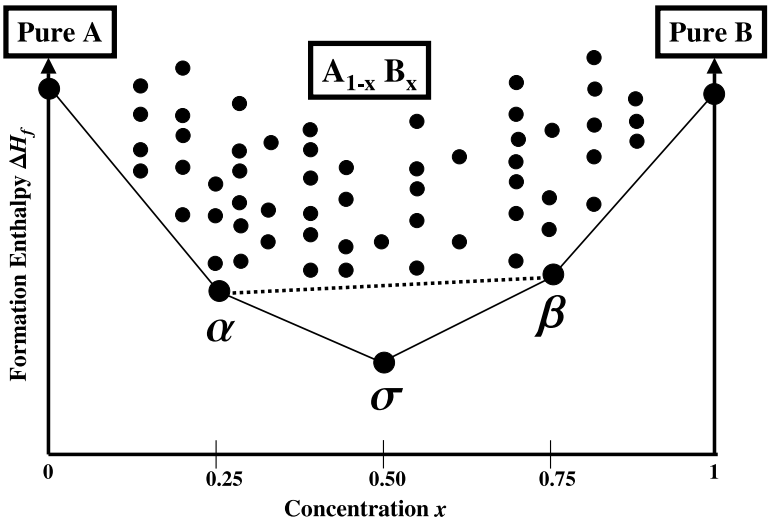


Figure 7. Schematic ground-state diagram of a binary alloy $A_{1-x}B_x$. The ground-state line was constructed from 60 energies of relaxed structures (given by dots) by use of Eq. (14). Besides the pure elemental crystal the ground-state line is formed by three structures α, σ, and β for concentrations $x = 0.25, 0.50$, and 0.75, respectively. If σ would lie energetically above the dashed tie line between α and β, a mixture of α and β would be more stable than σ.

(i) the maximum number of basis atoms up to which configurations are to be considered and (ii) the figure set chosen by a previous genetic algorithm run, along with the corresponding effective cluster interactions J. With this input, UNCLE automatically generates all possible superstructures (configurations) and determines their energy as predicted by the cluster expansion. The resulting ground state diagram and convex hull essentially constitute the $T = 0$ K phase diagram of the system.

We apply a new mathematical formalism to the cluster expansion that considerably simplifies aspects. Two places where this is particularly useful is in calculating the correlations (needed to perform the sum in Eq. (13)) and in Monte Carlo simulations. The new formalism works in the "space" of 3×3 *integer* matrices and provides an alternative representation for structures and figures.

Any supercell of the parent lattice is an integer multiple of the parent cell. So if the vectors of the parent lattice are the column vectors of a matrix A, there exists a matrix N, with all integer elements, such that $B = AN$. The columns of B are the lattice vectors of the supercell and the determinant of N will be the multiplicative factor; that is, if the supercell has twice the volume of the parent cell, then $|N| = 2$.

Because $B = AN$, the integer matrix N is an alternative representation for the superlattice. Realizing this, we can then map the superlattice and its atomic sites to this alternate representation, the g-representation, where the calculation of correlations is greatly simplified. In the g-representation, the atomic sites lie on an integer lattice, $\mathbb{Z}^3$, and the shape of the supercell is always orthorhombic. This simplifies the algorithm and thus makes the code much more efficient, both in time and memory.

Mapping to the g-representation is accomplished by decomposing N into its Smith

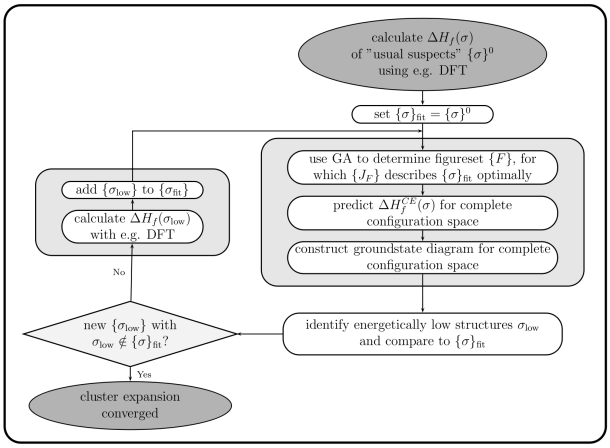


Figure 8. Illustration of the self-consistent "outer loop", which chooses the input structures of the cluster-expansion.

normal form (SNF). The SNF is a diagonal form with special properties (for details, see ref.[46,37]). and forms the key for efficient computation of correlation:. The lattice vectors and lattice points are represented by integers rather than floating point variables. No logic statements in the loops are required; no comparison of floating point numbers are needed. This improves both the efficiency and the robustness of the implementation.

Our implementation of UNCLE can be generalized to treat multinary systems. The treatment of ternary compounds has already been implemented and used. The extension beyond ternary systems is relatively simple and will be made as soon as required. To handle multinary expansions, the correlations must be calculated over a set of cluster functions. Formally there is also a set of cluster functions for a binary expansion, but there is only one function in the set and it can be taken to be the occupation itself, that is $\theta(s_i) = s_i$.

In the binary case, the correlation is computed merely by taking the product of each occupation value (± 1) over each vertex of a figure:

$$\Pi = \prod_{i=1}^{k} s_i \tag{15}$$

and there is one ECI, J, for each figure. But in the case of a n-ary system (n-components represented by n spin values), the complete description of the correlations requires $(n-1)$ cluster functions θ_l. Therefore, a figure with k vertices is no longer connected with a single correlation function, but instead $(n-1)^k$ correlation functions $\Pi^{(j)}$. The i^{th} entry of the superscript vector (j), which contains k entries, defines the cluster function θ_l, which is to be applied to the i^{th} vertex of the figure

$$\Pi^{(j)} = \prod_{i=1}^{k} \theta_{l=(j_i)}(s_i) \tag{16}$$

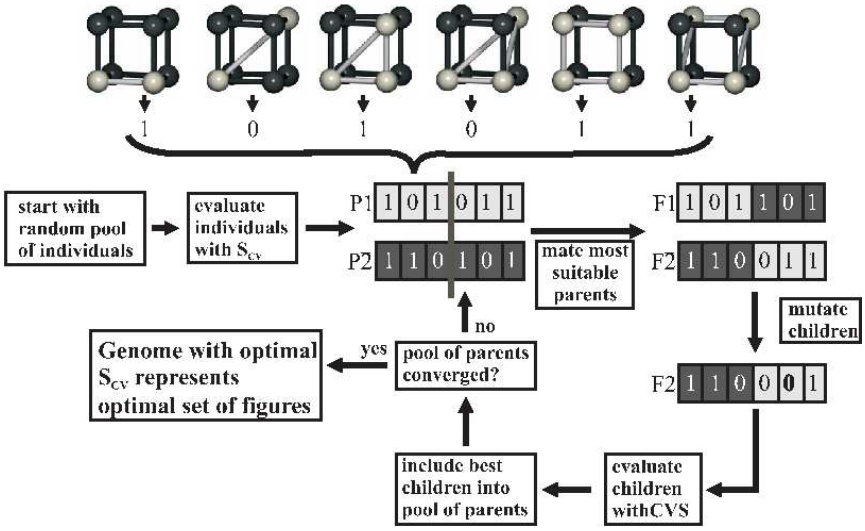


Figure 9. Illustration of the Genetic Algorithm, which helps to safely identify the relevant figures that need to be included in the CE-sum.

The full set of correlation functions of a figure consists of all the 2^k possible vectors (j). This number can be reduced according to the symmetry of a figure. The general multinary formalism was laid out by Sanchez et al. in[14], and later applied by Wolverton and de Fontaine[47]. Along the lines of the latter, we use Chebychev polynomials as cluster functions in the ternary case, an example for their application is given in section 3.2.

The cluster expansion approach is exact only when all possible figures are included in the cluster expansion sum, Eq. (13). But including such an (astronomic) number of terms in the expansion is impractical. To be useful, the expansion must be truncated to a relatively small number of terms without losing the expansion's predictive power. Choosing which figures to retain is the most critical step of the cluster expansion method. Nevertheless, finding a good selection of figures is a formidable task: There may be thousands of figures to choose from. Selecting a few dozen interactions from a pool of thousands is impossible to do exhaustively.

So far an evolutionary approach based on a genetic algorithm (GA) has proven to be the most effective way to choose the figures. The set of figures chosen by the GA results in a cluster expansion that has better predictive power than if chosen using other approaches. The details of the algorithm, which is implemented in UNCLE, have been described in[48,49]. Its basic principle is illustrated in Fig. 9.

The fitness criterion for the selection of figures is a *leave-many-out* cross validation score (see e.g.[50,51]). This fitness score S_{CV} is a measure of the predictive power for a given choice of figures. Its value is calculated by the following scheme:

1. Randomly choose $\mathcal{N}$ sets $\{\sigma\}^i_{\text{prediction}}$, $(i \in \{1...\mathcal{N}\})$ of n different structures out of the total pool of input structures.

2. For each of the $\mathcal{N}$ prediction-sets $\{\sigma\}^i_{\text{prediction}}$, perform a cluster expansion based

on all input structures except for those contained in $\{\sigma\}^i_{\text{prediction}}$. The resulting ECIs are not influenced by the energetics of $\{\sigma\}^i_{\text{prediction}}$.

3. Use the resulting ECIs to predict the energy of every member of $\{\sigma\}^i_{\text{prediction}}$ and compare it to the energy calculated by density functional theory.

4. Calculate the expectation value of the root-mean-square error for all the predicted structures:

$$S_{\text{CV}} = \sqrt{\frac{1}{\mathcal{N} \cdot n} \sum_{\mathcal{N}} \sum_{n} |E_{DFT}(\sigma) - E_{CE}(\sigma)|^2} \tag{17}$$

Other successful applications of the genetic algorithm within a cluster expansion can be found, e.g., in Ref.[52]. The GA has already been compared to the tetrahedron method proposed in[53] and the Variational Cluster Expansion[54,55], and proved to be the most reliable in finding the choice of figures with the best S_{CV}.

The determination of Effective Cluster Interactions (ECI) is performed as follows: For a given choice of figures and a set of $\mathcal{N}$ input structures $\{\sigma\}$, the effective cluster interactions J are extracted by minimizing[56]

$$\sum_{\mathcal{N}} \left(E_{DFT}(\sigma) - \sum_{F} D_F J_F \overline{\Pi}_F(\sigma) \right)^2 + \sum_{F} t_F J_F \stackrel{!}{=} \min \ , \tag{18}$$

where the last term is a *damping* term, which penalizes figures with large spatial extent (the spatial extent is determined as the average distance of the vertices from a figure's center of mass) $\mathbf{r_F}$:

$$t_F = c \cdot (\mathbf{r_F})^{\lambda} \tag{19}$$

The scaling variables c and λ are set independently for pair figures and higher-order figures. They are *not* chosen by the user, but optimized within the genetic algorithm.

For the fitting of the interactions according to equation 18, a set of constraints is introduced as proposed by Garbulsky and Ceder[57]. These constraints maintain the energetic hierarchy of the input structures within the hierarchy of the predicted energetics:

$$\Delta H^{\text{DFT}}(\sigma) - \delta_1(\sigma) < \Delta H^{CE}(\sigma) < \Delta H^{\text{DFT}}(\sigma) + \delta_1(\sigma) \tag{20}$$

$$\Delta H^{\text{DFT}}_{\text{GSL}}(\sigma) - \delta_2(\sigma) < \Delta H^{CE}_{\text{GSL}}(\sigma) < \Delta H^{\text{DFT}}_{\text{GSL}}(\sigma) + \delta_2(\sigma) \tag{21}$$

$$\Delta H^{\text{DFT}}_{\text{lowest}}(\sigma) - \delta_3(\sigma) < \Delta H^{CE}_{\text{lowest}}(\sigma) < \Delta H^{\text{DFT}}_{\text{lowest}}(\sigma) + \delta_3(\sigma) \tag{22}$$

The first constraint simply requires that the enthalpy $\Delta H(\sigma)$ of every structure σ, as calculated by DFT and predicted by the CE, matches within the error bars $\delta_1(\sigma)$. Independent error bars $\delta_i(\sigma)$ are set up for the energy distance of the enthalpy of a structure to the value of the ground state line at the respective concentration $\Delta H_{\text{GSL}}(\sigma)$, as well as for the energy distance between a structure's enthalpy and the enthalpy of the energetically lowest structure at this concentration $\Delta H_{\text{lowest}}(\sigma)$. For the actual fitting of Eq. (18) within the constraints of Eq. (20), an algorithm proposed by Goldfarb and Idnani[58] is implemented.

In some cases it may be more important to conserve the energy hierarchy for low-energy input-structures than for less stable structures. Thus, the error bars $\delta_i(\sigma)$ defined

in equation 20 depend on each structure's energy difference to the lowest structure at the respective concentration $\Delta H^{DFT}_{\text{lowest}}(\sigma)$, determined from first principles:

$$\delta_{\{1,2,3\}}(\sigma) = \delta^{\text{const}}_{\{1,2,3\}} \cdot \exp\left(-\frac{\Delta H^{\text{DFT}}_{\text{lowest}}(\sigma)}{k_B \cdot T}\right) \quad , \tag{23}$$

The constant part $\delta^{\text{const}}_{\{1,2,3\}}$ is specified at runtime. The Boltzmann-like energy-dependence can be varied through the term $k_B T$, and effectively turned off if desired.

While the fitting process is automatic, it introduces a set of new parameters for the fit itself (c and λ) as well as the Garbulsky-Ceder constraints. While the variables c and λ are optimized automatically within the genetic algorithm, $\delta_i(\sigma)$ and $k_B T$ have to be specified by the user. Nevertheless, it is simple to make sure that the constraints are set correctly by checking if the hierarchy predicted by the cluster expansion correctly reflects the hierarchy as determined by density functional theory. If this is not the case, then the constraints are lowered until the energetic hierarchy is preserved.

The selection and determination of the effective cluster interactions becomes challenging for low-symmetry systems such as surfaces. In the case of a surface, there is a loss of *translational* symmetry in one dimension. Consequently, the number of independent figures increases significantly because the ECIs become *layer dependent*. Compared to a bulk case, a larger number of input structures is necessary in order to determine the ECIs. However, it is possible to circumvent a part of this problem by treating the surface interactions as "correction" of the bulk interactions.

Because energies are additive, we may write

$$\Delta H^{CE}_f = \Delta H^{Vol}_f + \Delta H^{Surf}_f. \tag{24}$$

This ansatz was first applied by Drautz et al. to study the energetics of Ni-rich Ni-Al surfaces[59]. The advantage in treating the surface interactions as *correction* of the bulk interactions comes from the fact that the DFT calculations for different surface terminations and segregation profiles do not have to account for an infinite bulk reservoir. We only have to make sure that the DFT slab model is thick enough that the center layer of the slab is bulk-like. The energy of a structure σ can then be written as

$$E(\sigma) = \sum_{i=1}^{N} \left\{ \sum^{N_F} d_F \overline{\Pi}_F(\sigma) J_F + \sum^{N'_F} d'_F(\mathbf{R}_i) \overline{\Pi}'_F(\mathbf{R}_i) \delta J_F(\mathbf{R}_i) \right\}. \tag{25}$$

We see that for the surface part the interactions become site dependent. Here, $\mathbf{R}_i$ defines the position of the atom i with respect to the alloy surface. So, for an atom i within the segregation profile, every individual interaction J_F to neighboring atoms will be corrected to $J_F + \delta J_f(\mathbf{R}_i)$. Naturally, with increasing distance from the alloy surface, $\delta J_F \to 0$ and consequently the surface term (second term) in Eq. (25) becomes zero. In the case of e.g. a $Pt_{25}Rh_{75}(111)$ surface it turned out $\delta J_F \to 0$ already by the fourth layer[52,60].

In practice, for more complex surface problems, even this partition of the energy may be an insufficient strategy. In some cases, finding a sufficiently predictive set of ECIs may still require an unreasonably large number of DFT calculations. We are currently developing an additional concept to be implemented in UNCLE that will provide an improved reference energy as starting point for surface investigation. The mixed space cluster expansion[35,36] is applied to incorporate strain effects into the reference energy part. Next,

the energies of individual surface configurations are built from fully relaxed 1×1 surface structures, and, again, added to the reference energy part. We call this the concept of "structural bricks". After its implementation, it will be described in detail in Ref.[61].

There remains one critical point: As shown by Laks et al.[35], any CE in real space *fails* to predict the energy of long periodic coherent superlattices. For a given superlattice A_nB_n, Eq. (13) predicts a formation enthalpy $\Delta H_f = 0$ as $n \to \infty$. This indeed is an intrinsic fault of any finite CE and easy to understand: If we consider an A atom of an A_nB_n superlattice "far" away from the A/B interface so that all figures f connect the A atom exclusively to other A atoms, then the finite CE interprets the A atom as a bulk crystal atom and consequently, $\Delta H_f = 0$. However, as discussed in Section III.A, the formation enthalpy of an infinite superlattice should be defined as the equilibrium constituent strain energy, because in the limit $n \to \infty$ the superlattice formation enthalpy depends only on its strained constituents, and not on the interface properties. The problem can be solved[36,35] by transforming a group of interactions to the reciprocal space and adding the constituent strain term explicitely. This is easiest to do for the pair interactions. For this, we introduce the Fourier transform of real-space pair interactions, $J_{pair}(\mathbf{k})$ and the structure factor $S(\mathbf{k}, \sigma)$:

$$J_{pair}(\mathbf{k}) = \sum_j J_{pair}(\mathbf{R}_i - \mathbf{R}_j) \exp\left(-i\mathbf{k}\mathbf{R}_j\right) \tag{26}$$

$$S(\mathbf{k}, \sigma) = \sum_j S_j \exp\left(-i\mathbf{k}\mathbf{R}_j\right) \tag{27}$$

Then the formation enthalpies for any arbitrary atomically relaxed configuration σ are expressed by[36]

$$\Delta H_{CE}(\sigma) = \sum_{\mathbf{k}} J_{pair}(\mathbf{k}) |S(\mathbf{k}, \sigma)|^2 + \sum_F D_F J_F \bar{\Pi}_F(\sigma) + \Delta E_{CS}(\sigma). \tag{28}$$

This solution was introduced by Zunger and co-workers[36,35] and is called *Mixed-Space Cluster Expansion* (MSCE). The first term includes all pair figures in **k**-space. The second term represents many-body interactions and runs over symmetry inequivalent clusters consisting of three or more lattice sites. It also includes J_0 and J_1 from Eq. (11). D_F again stands for the number of equivalent clusters per lattice site, and $\bar{\Pi}_F(\sigma)$ are the structure-dependent geometrical coefficients given by Eq. (12). The last term represents the constituent strain energy of the structure σ, $\Delta E_{CS}(\sigma)$, and can be calculated by expanding the equilibrium constituent strain energy (Eq. (10)), $\Delta E^{eq}_{CS}(x, \hat{k})$, as[35,62]

$$\Delta E_{CS}(\sigma) = \sum_{\mathbf{k}} J_{CS}(x, \hat{k}) |S(\mathbf{k}, \sigma)|^2 \tag{29}$$

with

$$J_{CS}(x, \hat{k}) = \frac{\Delta E^{eq}_{CS}(x, \hat{k})}{4x(1-x)}. \tag{30}$$

Now, J_{CS} contains the correct long-periodic superlattice limit, namely the constituent strain energy [a].

[a] It has been found[62] that attenuating the constituent strain term can be important in strongly anharmonic, ordering

2.3 Extension to finite temperature and time-dependent properties

For finite temperature studies, Eq. (28) can be used in Monte-Carlo simulations. The code we applied for studying thermodynamic properties is a simple Metropolis algorithm[63] allowing for flipping pairs of A and B atoms in *arbitrary* distance mutual with the aim to reach the equilibrium configuration as fast as possible. The procedure is as follows:

1. Select randomly a pair of A and B atoms.
2. Calculate the energy difference δE caused by exchanging the two atoms. If $\delta E < 0$, flip the two spins; if $\delta E > 0$, flip the two spins with a probability of $\exp(-\delta E/kT)$ [again, E is obtained from Eq. (28)].
3. Go to 1.

Besides the temperature dependence of the alloy's free energy, MC simulations can be used to calculate coherent phase boundaries in the phase diagram. Following the fluctuation-response theorem[64] , the specific heat c_v of the system at a certain temperature can be calculated by the fact that c_v is proportional to the equilibrium fluctuations of the energy, $\langle E^2\rangle - \langle E\rangle^2$. Since the energy exhibits a point of inflection for a second-order phase transition at the transition temperature T_{trans}, its response function $c_v = (\partial E/\partial T)_v$ has a maximum at T_{trans} (Fig. (10)(a)). Although a phase transition is –strictly spoken– only defined for an *infinite* system, one usually also speak from a phase transition of a *finite* system, given by the maximum of c_v at the transition temperature as illustrated in Fig. (10)(a). If the MC simulations are applied for different concentrations x, the resulting T_{trans} values can be used to construct the coherent phase boundary of a system as displayed in Fig. (10)(b) for the Al-rich side of the Al-Cu phase diagram[65]. The open circles are measured values[66]. A small piece of the incoherent phase boundary is also shown. Yet, this boundary cannot be calculated by our method which is restricted to *coherent* alloy problems.

Another important application of MC simulations is the prediction the system's ordering. Of special interest are short-range order effects in disordered alloys which can quantitatively expressed in terms of SRO parameters as introduced in Section 1. For this, we rewrite Eq. (3) to the equivalent form

$$\alpha_{lmn}(x) = 1 - \frac{P_{lmn}^{A(B)}}{x} \tag{31}$$

where $P_{lmn}^{A(B)}$ is the conditional probability that given an A atom at the origin, there is a B atom at (lmn). For comparison with experimental data, the so-called "shells" lmn are introduced which are defined by the distance between A and B atoms in terms of half lattice parameters, $(l\frac{a}{2}, m\frac{a}{2}, n\frac{a}{2})$, e.g. for an fcc-lattice the nearest-neighbor distance would be described by the shell (110), the second neighbor distance by (200) and so on. As already mentioned, the sign of α indicates whether atoms in a given shell prefer to order ($\alpha < 0$) or cluster ($\alpha > 0$). The SRO parameter may be written in terms of the cluster expansion pair correlations as[32]

$$\alpha_{lmn}(x) = \frac{\langle \bar{\Pi}_{lmn}\rangle - q^2}{1 - q^2} \tag{32}$$

type systems. This is realized by an exponential damping function. However, since attenuating the constituent strain has no significant effect on the systems considered in this paper, this will be not discussed here.

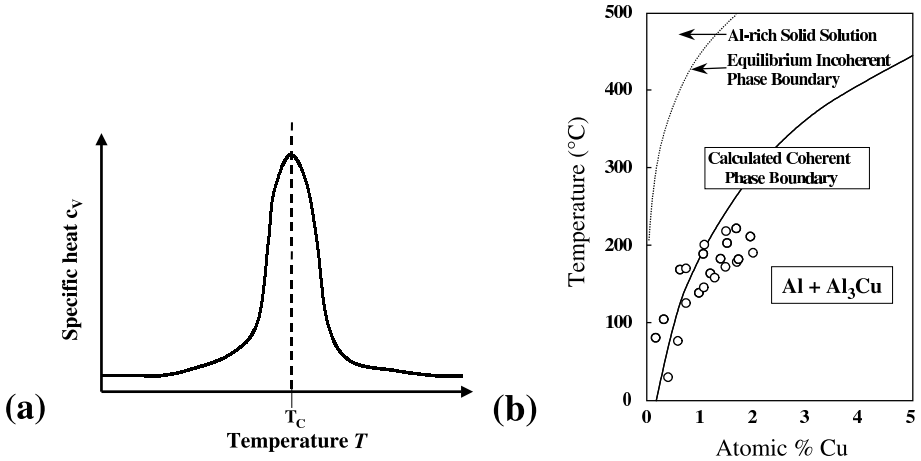


Figure 10. (a) Specific heat as function of temperature near a second-order phase-transition. c_v exhibits a maximum at T_{trans}. (b) Calculated coherent phase boundary for Al-rich Al-Cu and comparison to experimental data[66] (open circles).

where $q = 2x - 1$ and $\langle \bar{\Pi}_{lmn} \rangle$ is the pair correlation function, Eq. (12), for shell (lmn). In diffraction experiments the diffuse scattering due to SRO is proportional to the lattice Fourier transform of $\alpha_{lmn}(x)$[9,10]

$$\alpha(x, \mathbf{k}) = \sum_{lmn}^{n_R} \alpha_{lmn}(x) e^{i \cdot \mathbf{k} \cdot \mathbf{R}_{lmn}} \tag{33}$$

where n_R stands for the number of real space shells used in the transform. Equation (32) together with (33) opens the possibility to compare both, experimental and theoretically predicted diffuse diffraction patterns (reciprocal space) and SRO-parameters (real space). This concept will be applied in Section 3.1 to understand SRO phenomena in binary metal alloys *quantitatively*.

Similar to the calculation of the input structures' correlations for the cluster expansion, the determination of the starting energy of the Monte Carlo cell is done within the g-representation provided by the Smith normal form. The Monte Carlo cell is thus represented by the tensor G. Changing the atomic occupation of a site corresponds to changing the corresponding integer value of one element of G. In a Monte Carlo simulation, the calculation of the energy changes due to changes in the occupation (atom swaps) can be computed efficiently as only the energy contribution of those interactions "touched" by the swapped sites needs to be evaluated. The tensor G is the only large entity stored at runtime, requiring only *one byte* per site within the Monte Carlo cell; the correlations do *not* have to be stored at runtime. The minimal memory footprint allows for Monte Carlo cells of billions of sites, cpu time, rather than memory, becoming the limiting factor. A parallel implementation is planned to take advantage of this approach.

Besides the problem of bridging length scales, many materials properties require simulation times reaching from fractions of a second to weeks. One important example is the decomposition of an alloy into its constituents by precipitation. Precipitates represent

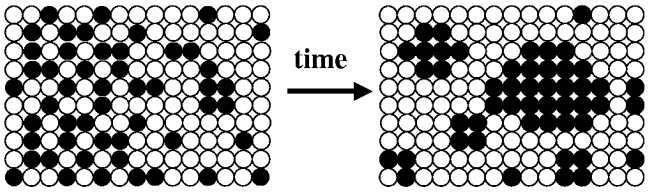


Figure 11. Schematic crystal-plane of an $A_{1-x}B_x$ alloy with characteristic islands formed by B (black) atoms during the aging process.

an important part of the microstructure of many alloy systems. Hereby, the dynamic evolution of precipitates takes place on a time scale of several hours, days or even months (see e.g.[69, 45]). The CE Hamiltonian can help to solve this second scaling problem, too, by using the effective interactions in Kinetic Monte-Carlo (KMC) simulations which is one of the most successful approaches to describe diffusion, growth and microstructure evolution in alloy systems[67]. The combination of CE and KMC simulations can be applied to simulate the aging of coherent precipitates in binary alloy systems. This decomposition reaction is sketched in fig. (11) by a simplified two-dimensional presentation: A quenched solid-solution (left frame) is aged at a given temperature. During this aging process islands are formed (right frame) which may show a characteristic size- and shape distribution (it is assumed that islands are formed by black B-atoms in an A-rich $A_{1-x}B_x$ alloy). The question is whether the distribution of these islands as a function of aging time can be calculated from first-principles.

The activations barrier for the exchange process can be expressed in terms of the temperature-dependent diffusion constant $D(T)$. In order to calculate $D(T)$ by a first-principles approach, it is assumed that the exchange of atoms is given by a vacancy-controlled diffusion. Therefore, in a first step, activation barriers must be calculated as a function of the structural environment. In the case of precipitation in which the alloy contains only a tiny amount (typically 1-5%) of the precipitating element, one often restricts the calculation of activation barriers to the case of the dilute limit (atom B in an A crystal) and the structural environment at the interface between solid-solution and precipitate. Although such activation barriers can -in principle- directly be used in KMC programs, they do not allow for a consideration of the temperature dependence as well as a transformation to real time scales. For this purpose, the complete phonon spectra for the relaxed structure corresponding to the vacancy formation, migration and the final configuration have to be calculated. This might be used in the framework of a transition state theory to predict the temperature dependent diffusion constant of the system, $D(T)$. Following classical diffusion theory the exchange frequency is proportional to the square of the atomic distance divided by the diffusion constant and the number of possible "jump directions" (e.g. six in a simple cubic lattice). If an exchange process between two certain neighbored atoms has been already chosen, then, consequently, the frequency $1/\tau_0$ for a chosen exchange process as a function of temperature T is connected to $D(T)$ by the relation

$$\tau_0(T) = \frac{a_{nn}^2}{D_{exp}(T)}, \tag{34}$$

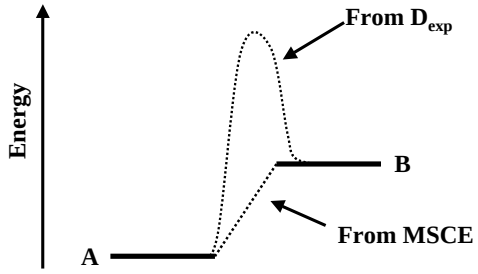


Figure 12. Basic assumption in our simulations[45]: While the energy difference between two neighboring atoms can be easily derived from the MSCE, an average and temperature dependent activation barrier is calculated from experimental temperature dependent diffusion data.

with a_{nn} being the average nearest-neighbor distance between atoms. Now, one can easily transform KMC steps to real time.

The strength of the CE to control a huge configuration space can now be utilized to calculate the energy difference for *all possible exchange processes* even, if there a millions of them. This allows to force atoms to move and to calculate the time which corresponds to this individual exchange process. The more unlikely an exchange process is, the longer is the corresponding time for this process. The concept is related to the "residence-time algorithm"[68] as discussed in chapter 12. for nearest-neighbor exchange processes only[69].

For the analysis of the shape of nanoclusters and precipitates, it is often helpful to apply the mixed-sace form of the cluster expansion (MSCE), because it allows for a quantitative separation of chemical and elastic energy parts[70]. Then, an accepted spin-flip would demand a recalculation of $S(\mathbf{k}, \sigma)$ in eqn. (28). However, as shown by Lu et al.[71], the MSCE method helps to avoid the necessity of recalculating $S(\mathbf{k}, \sigma)$ after each atomic movement by directly calculating the *change* in $J_{pair}(\mathbf{k})|S(\mathbf{k}, \sigma)|^2$ for each movement in real-space[71]. In the applied algorithm, a single KMC step is now *not longer a constant real time unit*, but depends on the corresponding probability W_{tot}. A single kinetic MC step corresponds indeed to only a single exchange of one B atom with one A atom and *not* to one trial-flip for each B atom. Since the "flip channel" i is always chosen randomly and usually a large number of B atoms (typically $10^3 - 10^5$) is considered to describe real aging processes, the probability that the same B atom is chosen in step i -when chosen already in step $(i-1)$- is extremely small. Due to the large system size it is not necessary to forbid certain exchanges between A and B atoms, i.e. we do not have to give up the restriction that the algorithm should be based on the Markovian process.

3 Applications

3.1 Ground-state search and short-range order

Our notions of the phase stability of compounds rest to a large extent on the experimentally assessed phase diagrams. Long ago, it was assumed that in the Cu-Pd system for

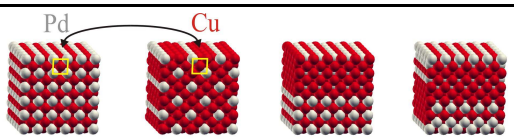
Name	$L1_2$	S1	S2 (LPS 3)	S3
Crystal structure				
	Cu_3Pd	Cu_7Pd	Cu_9Pd_3	Cu_8Pd_4
x_{Pd}	1/4	1/8	1/4	1/3
Lattice vectors	$\begin{pmatrix} 1&0&0\\ 0&1&0\\ 0&0&1 \end{pmatrix}$	$\begin{pmatrix} 1&0&0\\ 0&1&1\\ 0&\overline{1}&1 \end{pmatrix}$	$\begin{pmatrix} 1&0&0\\ 0&1&0\\ \frac{1}{2}&\frac{1}{2}&3 \end{pmatrix}$	$\begin{pmatrix} 1&0&0\\ 0&1&0\\ \frac{1}{2}&\frac{1}{2}&3 \end{pmatrix}$

Figure 13. [Color online] The ground state structures S1, S2 and S3, all related to $L1_2$ directly or to an $L1_2$-superstructure incorporating antiphase boundaries. These structures belong to the space group $P\frac{4}{m}mm$ (i.e. D^1_{4h} in Schoenflies nomenclature).

$x_{Pd} \leq 25\%$ there are at least two phases at high temperature ($L1_2$ and a $L1_2$-based superstructure), which evolve into a single, $L1_2$-ordered phase at low temperature. By constructing a first-principles Hamiltonian via the approach described above, a yet undiscovered Cu_7Pd ground state at $x_{Pd} = 12.5\%$ (referred to as S1 below) and an $L1_2$-like Cu_9Pd_3 superstructure at 25% (referred to as S2). We find that in the low-temperature regime, a single $L1_2$ phase cannot be stable, even with the addition of anti-sites. Instead we find that an S2-phase with S1-like ordering tendency will form. Previous short-range order diffraction data is quantitatively consistent with these new predictions (details can be found in ref.[72,73]). This study exemplifies how even well-established phase phenomena in classic alloy systems can be challenged via first principles statistical mechanics and calls for further experimental examination of this prototypical system.

Figure 14 shows the energies of $\approx 2^{20}$ ordered configurations and indicates the breaking points of the convex hull, i.e. the ground state structures. Figure 13 gives the structural description of the ground states. We find (a) The Cu_7Pd (S1) structure at $x_{Pd} = 12.5\%$, (b) the Cu_3Pd (S2 or LPS 3) structure at 25% and (c) the Cu_8Pd_4 (S3) structure at $x_{Pd} = 33\%$. We find that at $x_{Pd} = 25\%$ and $T = 0\,K$ S2 is considerably stabilized over $L1_2$ as ground state.

Finding (b) is in agreement with Refs.[74,75,71,76] ; S2 is predicted as a ground state at $x = 1/4$, lower in energy than $L1_2$: $\Delta H_f(S2) = -102.6\,meV/atom$, $\Delta H_f(L1_2) = -99.8\,meV/atom$. At 12.5% Lu *et al.*[77] predicted the D1 structure, which, though not identical to S1, is also similar to $L1_2$. The S1 ground state is related to the $L1_2$ structure by a simple exchange of Cu and Pd rows along $[100]$ as shown in Fig. 13. Previous studies (e.g.[78–80]) that obtained $L1_2$ as ground state at $\approx 18\%$ referred to the ANNNI Ising model, or performed an electronic mean field approach[74]. However, negligence of S1 in the first-principles input (Ref.[76]) will favour interactions that are "blind" for S1.

Given that we predict at an S1 phase $T = 0\,K$ at 12.5% Pd and an S2 phase at 25% Pd, it is interesting to characterize the phase(s) at intermediate concentrations. In order to examine the energies $E_{CE}(\sigma)$ of structures with cells bigger than 20 atoms (Fig. 3), we

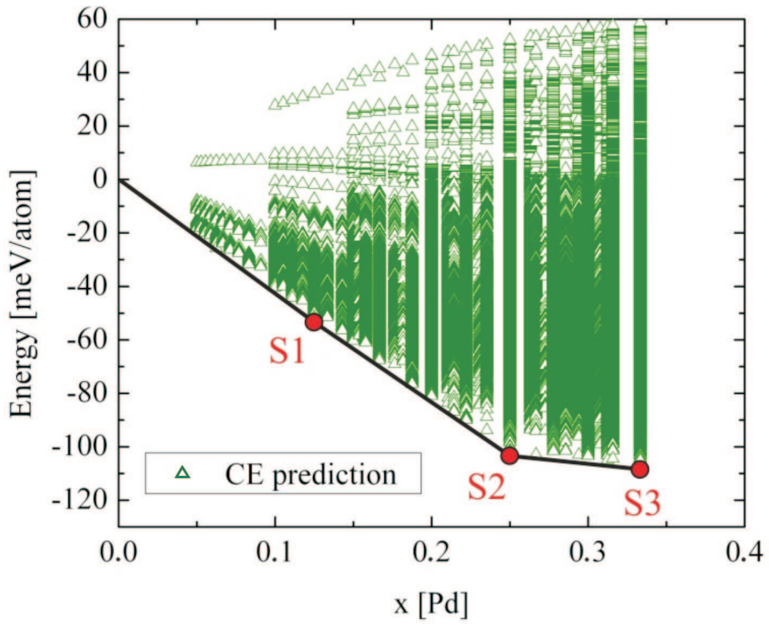


Figure 14. [Color online] Ground state diagram in the Cu-rich regime. Each triangle represents the predicted energy for one specific crystal structure. The solid line is the convex hull of all $\mathcal{O}(10^6)$ energies. The ground state structures are depicted in Fig. 13. $L1_2$ is not a ground state, but rather the $L1_2$-related superstructure S2 (LPS 3).

constructed large 24×24×24 cells and sample their energies via Monte Carlo (vibrational entropy was not taken into account). Due to the variety of incommensurate superstructures with non-coherent phase boundaries, we have to restrict our study to low temperatures [b]—a more thorough thermodynamic study may not be feasible with Monte Carlo. Nevertheless, the critical temperature $T_c \approx 800\,\mathrm{K}$ for the phase transition from A1 to S2 is in good agreement with experiment ($T_c^{\mathrm{Exp.}} \approx 780\,\mathrm{K}$). Simulated annealing in the intermediate region provides indication of a transition from the disordered high temperature phase to a lower temperature S1-like S2 structure. The latter resembles LPS 3-like ordering, permeated with an S1-like pattern [c]. An investigation of the energetic hierarchies of these phases supports the hypothesis of the formation of an S1-like S2 structure.

Unfortunately, in the S1-like S2 region, no recent diffraction data are available in order to directly compare experimental with our calculated results, hence we examine SRO data from the region of coherency. In Fig. 15 we show our calculations of the SRO parameters α_{lmn} for 29.8% Pd, where several studies yielded comparable data.

The study above is a characteristic example, how ab-initio based studies can help to clarify uncertain, low-temperature regions in alloy phase diagrams: Contrary to previous

[b] Incoherencies, originating from smoothened APB profiles and wetting around the phase transition cannot be accounted for by our MC simulation, which is restricted to the fcc lattice.

[c] Narrow regions of two-phase coexistence could not be captured by our MC simulation. However, such two-phase regions, even if very narrow, due to Gibbs' phase rule.

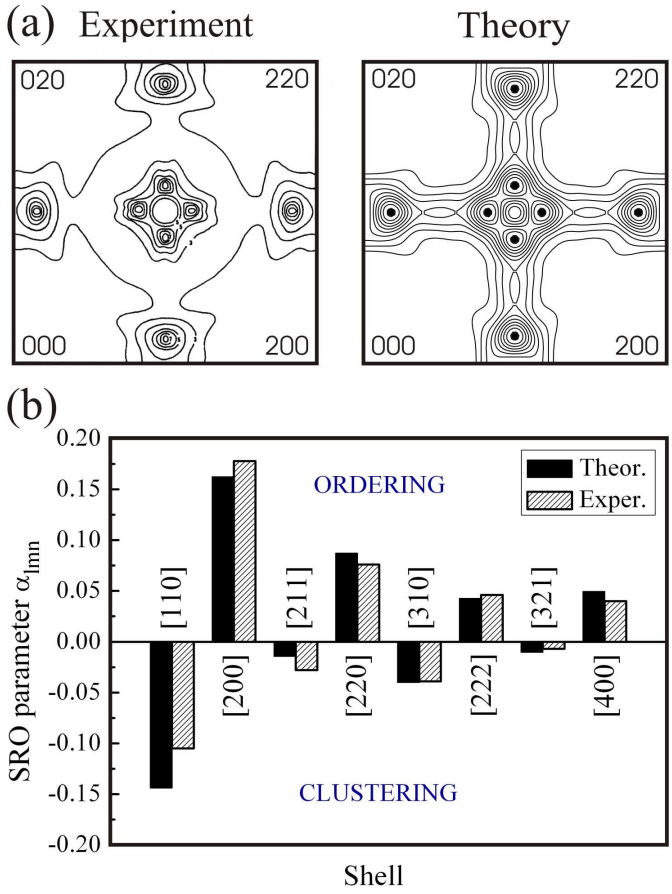


Figure 15. (a) Experimental[81] vs. theoretical short-range order for $Cu_{0.702}Pd_{0.298}$ at T=773 K in reciprocal space. The SRO exhibits peaks of the fundamental wave vector $\mathbf{k} = (1, 1/2M, 0)$ at $M = 3$, in excellent agreement to the superstructure period of S2. (b) Real space SRO for neighboring pairs separated by [hkl].

assessments, Cu-Pd does not have an $L1_2$ ground state, but the Cu_3Pd S2 structure is more stable at 25% composition. Furthermore, a new ground state S1 is predicted at lower composition with Cu_7Pd stoichiometry, hence the features of $L1_2$-like ordering observed experimentally at 17%[78] are due to a S2 with S1-like defects, not due to an $L1_2$ phase.

3.2 Point-defects at grain boundaries

For the Ni-Al system, it is well known that the ordering of defects plays a fundmental role. Understanding the defect structure and stability within the NiAl B2 phase and the Ni_2Al_3 phase is key to understanding the system. In the sense of the cluster expansion lattice, both phases can be described as bcc-based superstructures. It is well known[82–85] that on the simple cubic Ni sublattice of the B2 NiAl phase, vacancies are the dominant defect type in Ni-poor NiAl. Also, if Ni_2Al_3 is to be described as a decoration of Ni and Al atoms on

a bcc lattice, then $\frac{1}{6}$ of the lattice sites are left vacant. Therefore, in order to study defect order with Ni_xAl_{1-x} in the concentration range $0.4 \leq x \leq 0.5$, the cluster expansion needs to explicitly treat vacancies as a third component.

In order to obtain a converged cluster expansion for this system, 129 structures were calculated using VASP. Based on a total number of 711 figures with up to six vertices, the genetic algorithm chose a set of figures with a total of 82 ECI J_F. Two hundred prediction sets, each with $n = 10$ predicted structures, were used to compute the cross-validation-score, resulting in a CV score of $S_{\text{CV}} = 6.0$ meV.

Figure 16 shows the resulting ground state diagram as predicted by UNCLE. The ground state diagram shown in Fig. 16 has been limited to $Ni_xAl_{(1-x)}$ concentrations $0.4 < x < 0.\overline{6}$, as this is the only concentration regime, whithin which bcc-based superstructes are observed experimentally[82,86]. Furthermore this investigation exclusively focused on the description of point defects *within* this concentration regime, which is why the cluster expansion only required convergence for this concentration range. The configuration space search included all ternary bcc-superstructure with 16 sites or less and with less than 21% vacancies. The number of unique configurations is nearly 13 million. To compute these energies of all these configurations using UNCLE requires less than 36 hours on a single 2.8 GHz processor.

Each "□" in Fig. 16 indicates the enthalpy of a structure that was calculated by DFT and included in the cluster expansion to extract the ECIs. Every "+" in the figure corresponds to the cluster expansion prediction for one atomic configuration. Consistent with the observed phase diagram, Ni_2Al_3 and B2 are predicted to be stable at $x = 0.4$ and $x = 0.5$ respectively. The third stable ground state within the converged part of the cluster expansion is Ni_2Al at $x = 0.\overline{6}$, which can be observed experimentally to be a metastable

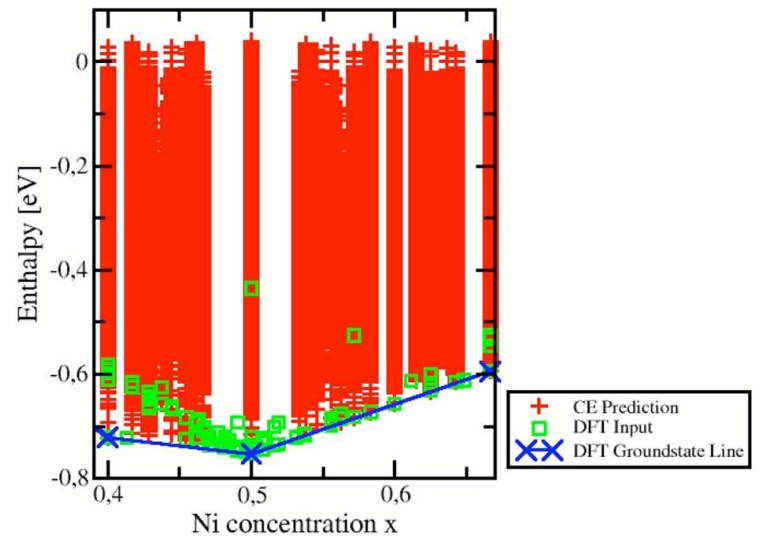


Figure 16. Calculated ground state diagram of all bcc-based superstructures with up to 16 sites occupied by Ni, Al and less than 21% vacancies.

state[86].

The cluster expansion Hamiltonian corresponding to the ground state diagram of Fig. 16 can also be applied to study the defect order at finite temperature. More than the ground state search (which holds no surprises), it is in this context that the cluster expansion is useful for the Ni-Al system. As one example, Fig. 17 provides a view into the ordering of B2-NiAl for $T \approx 4900$K (left) and room temperature (right) resulting from Monte Carlo modelling. The (100) plane shown in the figure is one layer of a Monte Carlo cell consisting of one million lattice sites. The concentration of the three constituents have been fixed to 50% Ni, 45% Al and 5% vacancies.

In full global thermodynamic equilibrium, a Monte-Carlo cell containing 50% of both Ni and Al should exhibit a single B2-ordered domain. Thus a cut along a (100) plane would only contain either Ni- or Al-atoms, depending on whether it lies within the Ni- or the Al-sublattice. Fig. 17 shows that the (100)-plane consists of both Ni- and Al-domains. These regions of Ni and Al belong to different B2-domains, which coexist within the Monte Carlo cell. By changing the external parameters of the simulation, the Monte-Carlo cell can be brought into thermal equilibrium and the different domains visible in Fig. 17 merge into a single B2-domain. While strongly increasing the required calculation time, this does not add any new scientific insight, as the presence of the anti-phase boundaries between different B2 domains does not interfere with the observation of short-ranged vacancy order in the bulk of the respective B2-domains. The important point is to note, that Fig. 17 shows B2-domains with a stacking fault in between them and *not* domains of pure Ni or Al. This allows us to observe Al-subplanes of the B2-structure (light-grey domain) and the Ni-subplanes (dark-grey domains) within a single cut along (100).

For the high temperature case the formation of different B2-domains (dark and light gray) on the lattice can already be observed. The vacancies (white) occupy nearly random sites within Ni- and Al-subplanes. At room temperature the formation of the different B2-domains is complete and the vacancies form diagonal chains within the Ni-subplanes of

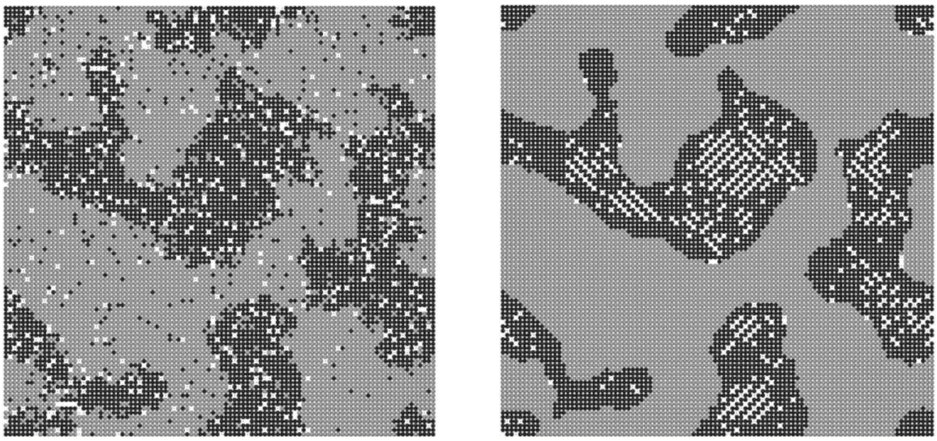


Figure 17. B2-NiAl with 5% vacancies: Cut along the (100) orientation through a $100 \times 100 \times 100$ Monte-Carlo cell for $T = 4936$ K (left) and $T = 294$ K. It can be seen that for lower temperatures all vacancies (white) short range order in the Ni-domains (black).

the B2-phase only. These chains can be interpreted a starting growth of the Ni_2Al_3 phase, where the vacancies are ordered in the same way. Thereby such simulations allow for a quantitative analysis of the phase stability of these alloy phases. A detailed interpretation and evaluation of the structural properties can be found in reference[87].

3.3 Surface Segregation

As known from experimental studies on $Pt_{25}Rh_{75}(111)$[88,89], this surface possesses a characteristic segregation profile: While the top layer shows a Pt enrichment, Pt depletion is found for the layer underneath. The existence of an equilibrium segregation profile is manifested by chemically resolved STM images[88], Low Energy Ion Scattering (LEIS) and quantitative Low Enery Electron Diffraction (LEED) analyses[89]. These studies unambiguously show that for annealing temperatures above $\sim$ 1000 K the observed segregation profile does not longer depend on the experimentally chosen annealing temperature of the sample.

Considering the energetics of the alloy system Pt-Rh, the pronounced segregation profile appears to be a surprise, because formation enthalpies of intermetallic compounds are all between 0 and about -20 meV/atom, i.e. smaller than kT at room temperature. This is in agreement with the bulk phase diagram of this binary system which does not show any long-range ordered structures in the experimentally accessible temperature regime. Instead, a fcc-based solid solution is stable for all concentrations. As a consequence of this small heterogeneous bonding, all constructed effective cluster interactions J_F for bulk and surface are unusually small, possessing energy values much smaller than 20 meV per atom, and cannot explain the characteristic segregation profile found for the (111) surface. However, there is one relevant deviation between the energetic properties of the bulk and the surface: Due to the symmetry break the onsite energies of individual atomic sites which are defined by J_0 and J_1 in Eq.(11) are different for the near-surface layers compared to the bulk. For only weakly ordering systems as the $Pt_{25}Rh_{75}$(111) surface these onsite energies represent a good measure for the segregation behaviour. Actually, it turns out that the top layer shows a tremendous tendency for an enrichment with Pt atoms reflected by an energy gain of about 0.2 eV per atom! Interestingly the opposite is true for the layer underneath: Here, the onsite energy speaks for a Pt depletion and clustering of Rh atoms.

In order to predict the segregation profile quantitatively, Monte-Carlo simulations were performed. As displayed in Fig. 18(left) our constructed cluster expansion is able to reproduce the experimental segregation profile determined via quantitative LEED analysis[89]. It turns out that for this surface system already a 40×40 atom cell per layer was sufficient for a quantitative description of the segregation profile as well as the substitutional ordering. For the latter, fig. 18(right) compares an STM image with atomic and chemical resolution[88] with our predicited one. It can be seen that there is an excellent (quantitative) agreement between experiment and theory.

4 Concluding Remarks

With the program package UNCLE, we present a tool that makes the cluster expansion more accessible to non-specialists and applicable to a wide variety of physical problems. Several extensions of the formalism were presented: Use of the g-representation simplifies

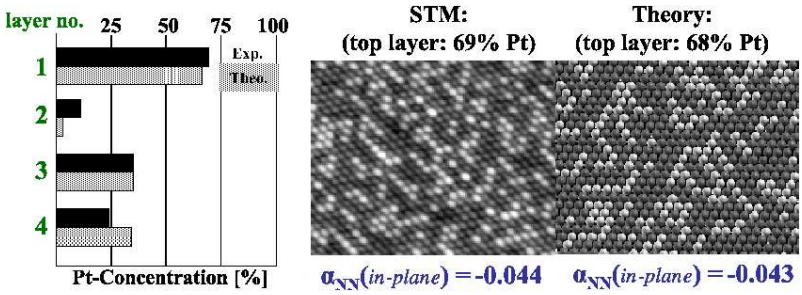


Figure 18. Left: Experimentally determined and predicted segregation profile for $Pt_{25}Rh_{75}(100)$ ($T_{anneal} = 1400$K). Right: Corresponding short range-order behaviour as found by STM and predicted by our CE approach.

and automates the "chores" of setting up and constructing a cluster expansion, performing ground state searches, and using the ECIs in Monte Carlo simulations. By automating much of the cluster expansion construction and use, problems arising from user errors are less likely, resulting in more robust predictions. The treatment of surface alloys and related systems is made possible through the separation of the cluster expansion Hamiltonian into a bulk and surface part.

Since the approach used is only a few years old, its application potential is by no means already reached. There are a number of solid properties which may be treated via DFT, CE, and MC after some further development, as e.g. nucleation processes or cluster from the gas phase. Since the approach allows to describe the behaviour of real alloy systems, a strong interplay with experimental groups is highly desirable.

Acknowledgments

The author gratefully acknowledges support by DFG, MU 1648/2 and Mu 1648/3 (Germany), Studienstiftung des deutschen Volkes, and by the NSF, DMR 0650406.

References

1. P. Villars and L. D. Calvert, *Pearson's Handbook of Crystallographic Data*, ASM International, Materials Park, 1991.
2. L. Reinhard, B. Schönfeld, G. Kostorz, and W. Bührer, Phys. Rev. B, **41**, 1727, 1990.
3. S. Müller and A. Zunger, Phys. Rev. B **63**, 094294, 2001.
4. R. Hultgren, P. D. Desai, D. T. Hawkins, M. Gleiser, K. K. Kelley, *Selected Values of the Thermodynamic Properties of Binary Alloys*, American Society for Metals, Ohio, 1973.
5. J. M. Ziman, *Models of Disorder*, Cambridge University Press, Cambridge, 1979.
6. T. Muto and Y. Tagaki, *Solid State Physics*, vol. 1, p. 193, Academic Press, 1955.
7. L. Guttman, *Solid State Physics*, vol. 3, p. 145, Academic Press, 1956.
8. J. M. Cowley, J. Appl. Phys., **21**, 24, 1950.
9. M. A. Krivoglaz, *X-Ray and Neutron Diffraction in Nonideal Crystals*, Springer, Berlin, 1996.

10. C. J. Sparks and B. Borie, “Local arrangements studied bx x-ray diffraction”, 1966, Met. Soc. Conf.
11. B. Schönfeld, Prog. Mat. Sci., **44**, 435, 1999.
12. P. Hohenberg and W. Kohn, Phys. Rev. **136**, 864B, 1964.
13. W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133A, 1965.
14. J. M. Sanchez, F. Ducastelle, and D. Gratias, Physica A, **128**, 334, 1984.
15. M. C. Payne, M. P. Teter, D. C. Allen, T. A. Arias, J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1064, 1992.
16. R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689, 1989.
17. R. M. Dreizler, E. K. U. Gross, *Density Functional Theory*, Springer, Berlin, 1990.
18. R. M. Dreizler and J. da Providencia, *Density Functional Theory*, Plenum-Press, New York, 1985.
19. E. S. Krachko and E. V. Ludena, *Energy Density Functional Theory of Many electron Systems*, Kluwer Academic, Boston, 1990.
20. D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 567, 1980.
21. J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048, 1981.
22. J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244, 1992.
23. D. J. Singh, *Planewaves, Pseudopotentials and the LAPW Method*, Kluwer, Boston, 1994.
24. D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494, 1979.
25. N. Troullier and J. L. Martins, Phys. Rev. B **43** 1993, 1991.
26. D. Vanderbilt, Phys. Rev. B **41**, 7892, 1990.
27. G. Kresse and J. Hafner, J. Phys.: Condens. Matter, **6**, 8245, 1994.
28. G. Kresse and D. Joubert, Phys. Rev. B, **59**, no. 3, 1758–1775, Jan 1999.
29. P. E. Blöchl, Phys. Rev. B **50**, 17953, 1994.
30. A. Zunger, in: Handbook of Crystal Growth, T.D.J. Hurle, (Ed.), vol. 63, p. 99, Elsevier, Amsterdam. 1994, and references therein.
31. D. M. Wood and A. Zunger, Phys. Rev. Lett., **61**, 1501, 1988.
32. V. Ozoliņš, C. Wolverton, and A. Zunger, Phys. Rev. B, **57**, 4816, 1998.
33. V. Ozoliņš, C. Wolverton, and A. Zunger, Phys. Rev. B, **57**, 6427, 1998.
34. D. J. Bottomeley and P. Fons, J. Cryst. Growth, **44**, 513, 1978.
35. D. B. Laks, L. G. Ferreira, S. Froyen, and A. Zunger, Phys. Rev. B, **46**, 12587, 1992.
36. A. Zunger, in: NATO ASI on: Statics and Dynamics of Alloy Phase Transformations, P. E. A. Turchi and A. Gonis, (Eds.), p. 361, Plenum Press, New York. 1994.
37. D. Lerch, O. Wieckhorst, G. L. W. Hart, R. W. Forcade, and S. Müller, submitted to Modelling Simul. Mater. Sci. Eng.
38. Georg Kresse and Jürgen Hafner, Phys. Rev. B, **47**, 558, 1993.
39. Georg Kresse and J. Furthmüller, Phys. Rev. B, **54**, 11169, 1996.
40. Georg Kresse and J. Furthmüller, Comp. Mat. Sci, **6**, 15, 1996.
41. E. Wimmer, H. Krakauer, M. Weinert, and A. J. Freeman, Phys. Rev. B, **24**, no. 2, 864–875, Jul 1981.
42. M. Weinert, J. Math. Phys., **22**, no. 11, 2433, Nov 1981.
43. M. Weinert, E. Wimmer, and A. J. Freeman, Phys. Rev. B, **26**, no. 8, 4571–4578, Oct 1982.
44. L. G. Ferreira, S.-H. Wei, and A. Zunger, J. Supercomp. Appl., **5**, 34, 1991.
45. S. Müller, J. Phys.: Cond. Matter, **15**, 2003.

46. Gus L. W. Hart and Rodney W. Forcade, Phys. Rev. B, **77**, no. 22, 224115, 2008.
47. C. Wolverton and D. de Fontaine, Phys. Rev. B, **49**, no. 13, 8627, 1994.
48. Volker Blum, Gus L. W. Hart, Michael J. Walorski, and Alex Zunger, Phys. Rev. B, **72**, no. 16, 165113, 2005.
49. G. L. W. Hart, V. Blum, J. Walorski, and A. Zunger, Nature Materials, **4**, no. 5, 391, 2005.
50. A. van de Walle and G. Ceder, Journal of Phase Equilibria, **23**, 348, 2002.
51. K. Baumann, Trends in Analytical Chemistry, **22**, 395, 2003.
52. S. Müller, M. Stöhr, and O. Wieckhorst, Applied Physics A, **82**, no. 3, 415, 2005.
53. Nikolai A. Zarkevich and D. D. Johnson, Phys. Rev. Lett., **92**, no. 25, 255702, Jun 2004.
54. Alejandro Diaz-Ortiz and Helmut Dosch, Phys. Rev. B, **76**, no. 1, 012202, 2007.
55. Alejandro Díaz-Ortiz, Helmut Dosch, and Ralf Drautz, Journal of Physics: Condensed Matter, **19**, no. 40, 406206, 2007.
56. J. W. D. Conolly and A. R. Williams, Phys. Rev. B, **27**, 5169, 1983.
57. G. D. Garbulsky and G. Ceder, Phys. Rev. B, **49**, no. 9, 6327–6330, Mar 1994.
58. D. Goldfarb and A. Idnani, Math. Prog., **27**, 1, 1983.
59. R. Drautz, H. Reichert, M. Fähnle, H. Dosch, and J. M. Sanchez, Phys. Rev. Lett., **87**, no. 23, 236102, Nov 2001.
60. S. Müller, Surface and Interface Analysis, **38**, 1158, 2006.
61. T. Kerscher, O. Wieckhorst, and S. Müller, to be submitted.
62. C. Wolverton, V. Ozoliņš, and A. Zunger, J. Phys.: Condens. Matter, **12**, 2749, 2000.
63. N. Metropolis, A. W. Rosenbluth, M. V. Rosenbluth, A. Teller, and E. Teller, J. Chem Phys., **60**, 1071, 1974.
64. M. Toda, R. Rubo, and N. Saito, *Statistical Physics I*, Springer, Berlin, 1983.
65. S. Müller, L.-W. Wang, A. Zunger, and C. Wolverton, Phys. Rev. B, **60**, 16448, 1999.
66. J. L. Murray, Int. Met. Rev., **30**, 211, 1985.
67. J. Jacobsen, K. W. Jacobsen, and P. Stoltze and. J. K. Norskov, Phys. Rev. Lett., **74**, 2295, 1995.
68. A. B. Börtz, M. H. Kalos, and J. L. Lebowitz, J. Comp. Phys., **17**, 10, 1975.
69. S. Müller, L.-W. Wang, and A. Zunger A., Model. Simul. Mater. Sci. Eng., **10**, 131, 2002.
70. S. Müller, C. Wolverton, L.-W. Wang, and A. Zunger, Acta Mater. **48**, 4007, 2000.
71. Z. W. Lu, D. B. Laks, S. H. Wei, and A. Zunger, Phys. Rev. B, **50**, 6642, 1994.
72. S. Bärthlein, G. L. W. Hart, A. Zunger, and S. Müller, J. Phys.: Condens. Matter **19**, 032201, 2007.
73. S. Bärthlein, E. Winning, G. L. W. Hart, and S. Müller, Acta Mater., in press.
74. G. Ceder, D. de Fontaine, H. Dreysse, D. M. Nicholson, G. M. Stocks, and Gyorffy B. L., Acta metall. mater., **38**, 2299, 1990.
75. A. V. Ruban, S. Shallcross, S. I. Simak, and H. L. Skriver, Phys. Rev. B, **70**, 125115, 2004.
76. C. Colinet and A. Pasturel, Phil. Mag. B, **82**, 1067, 2002.
77. Z. W. Lu, S.-H. Wei, Alex Zunger, S. Frota-Pessoa, and L. G. Ferreira, Phys. Rev. B, **44**, no. 2, 512–544, Jul 1991.
78. D. Broddin, G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, R. Portier, M. Guymont, and A. Loiseau, Phil. Mag. A, **54**, 395, 1986.

79. S. Takeda, J. Kulik, and D. de Fontaine, Acta metall., **35**, 2243, 1987.
80. S. Takeda, J. Kulik, and D. de Fontaine, J. Phys. F, **18**, 1387, 1988.
81. Kenichi Oshima and Denjiro Watanabe, Acta Cryst. A, **32**, 883, 1976.
82. A. Taylor and N. J. Doyle, Journal of Applied Crystallography, **5**, no. 3, 201–209, Jun 1972.
83. C. L. Fu, Y.-Y. Ye, M. H. Yoo, and K. M. Ho, Phys. Rev. B, **48**, no. 9, 6712–6715, Sep 1993.
84. B. Meyer and M. Fähnle, Phys. Rev. B, **59**, no. 9, 6072–6082, Mar 1999.
85. F. Lechermann and M. Fähnle, Phys. Rev. B, **63**, no. 1, 012104, Dec 2000.
86. F. Reyaud, Journal of Applied Crystallography, **9**, no. 4, 263–268, Aug 1976.
87. D. Lerch and S. Müller, to be submitted.
88. E.L.D. Hebenstreit, M. Hebenstreit, M. Schmid, and P. Varga, Surface Science, **441**, 441, 1999.
89. E. Platzgummer, M. Sporn, R. Koller, S. Forsthuber, M. Schmid, W. Hofer, and P. Varga, Surface Science, p. 236, 1999.

Large Spatiotemporal-Scale Material Simulations on Petaflops Computers

Ken-ichi Nomura[1], **Weiqiang Wang**[1], **Rajiv K. Kalia**[1], **Aiichiro Nakano**[1], **Priya Vashishta**[1], **and Fuyuki Shimojo**[2]

[1] Collaboratory for Advanced Computing and Simulations
Department of Computer Science
Department of Physics and Astronomy
Department of Chemical Engineering and Material Science
University of Southern California, Los Angeles, CA 90089-0242, USA
E-mail: (knomura, wangweiq, rkalia, anakano, priyav)@usc.edu
[2] Department of Physics, Kumamoto University, Kumamoto 860-8555, Japan
E-mail: shimojo@kumamoto-u.ac.jp

We have developed a parallel computing framework for large spatiotemporal-scale atomistic simulations of materials, which is expected to scale on emerging multipetaflops architectures. The framework consists of: (1) an embedded divide-and-conquer (EDC) framework to design linear-scaling algorithms for high complexity problems; (2) a space-time-ensemble parallel (STEP) approach to predict long-time dynamics while introducing multiple parallelization axes; and (3) a tunable hierarchical cellular decomposition (HCD) parallelization framework to map these $O(N)$ algorithms onto a multicore cluster. The EDC-STEP-HCD framework has achieved: (1) inter-node parallel efficiency well over 0.95 for 218 billion-atom molecular-dynamics (MD) and 1.68 trillion electronic-degrees-of-freedom density functional theory-based quantum-mechanical simulations on 212,992 IBM BlueGene/L processors; (2) high intra-node, multithreading and single-instruction multiple-data parallel efficiency; and (3) nearly perfect time/ensemble parallel efficiency. The spatiotemporal scale covered by MD simulation on a sustained petaflops computer per day (i.e. petaflops•day of computing) is estimated as NT = 2.14 (e.g. N = 2.14 million atoms for T = 1 microseconds). Results of multimillion-atom reactive MD simulations on nano-mechano-chemistry reveal various atomistic mechanisms for enhanced reaction in nanoenergetic materials: (1) a concerted metal-oxygen flip mechanism at the metal/oxide interface in thermites; (2) a crossover of oxidation mechanisms in passivated aluminum nanoparticles from thermal diffusion to ballistic transport at elevated temperatures; and (3) nanojet-catalyzed reactions in a defected energetic crystal.

1 Introduction

Fundamental understanding of complex system-level dynamics of many-atom systems is hindered by the lack of validated simulation methods to describe large spatiotemporal-scale atomistic processes. The ever-increasing capability of high-end computing platforms is enabling unprecedented scales of first-principles based simulations to predict system-level behavior of complex systems.[1] An example is large-scale molecular-dynamics (MD) simulation involving multibillion atoms, in which interatomic forces are computed quantum mechanically to accurately describe chemical reactions.[2] Such simulations can couple chemical reactions at the atomistic scale and mechanical processes at the mesoscopic scale to solve broad mechano-chemistry problems such as nanoenergetic reactions, in which reactive nanojets catalyze chemical reactions that do not occur otherwise.[3] An even harder problem is to predict long-time dynamics, because the sequential bottleneck of time precludes efficient parallelization.[4,5]

The hardware environment is becoming challenging as well. Emerging sustained petaflops computers involve multicore processors,[6] while the computer industry is facing a historical shift, in which Moore's law due to ever increasing clock speeds has been subsumed by increasing numbers of cores in microchips.[7] The multicore revolution will mark the end of the free-ride era (i.e., legacy software will run faster on newer chips), resulting in a dichotomy—subsiding speedup of conventional software and exponential speedup of scalable parallel applications.

To address these challenges, we have developed key technologies for parallel computing with portable scalability. These include an embedded divide-and-conquer (EDC) algorithmic framework to design linear-scaling algorithms for broad scientific and engineering applications (e.g. equation solvers, constrained optimization, search, visualization, and graphs involving massive data) based on spatial locality principles.[8] This, combined with a space-time-ensemble parallel (STEP) approach[9] to predict long-time dynamics based on temporal locality[10] and a tunable hierarchical cellular decomposition (HCD) parallelization framework, maximally exposes concurrency and data locality, thereby achieving reusable "design once, scale on new architectures" (or metascalable) applications.[11,12] It is expected that such metascalable algorithms will continue to scale on future multicore architectures. The "seven dwarfs" (a dwarf is an algorithmic method that captures a pattern of computation and communication) have been used widely to develop scalable parallel programming models and architectures.[6] We expect that the EDC-STEP-HCD framework will serve as a "metascalable dwarf" to represent broad large-scale scientific and engineering applications.[12]

We apply the EDC-STEP-HCD framework to a hierarchy of atomistic simulation methods. In MD simulation, the system is represented by a set of N point atoms whose trajectories are followed to study material properties.[4,13,14] Quantum mechanical (QM) simulation further treats electronic wave functions explicitly to describe chemical reactions.[15–17] To seamlessly couple MD and QM simulations, we have found it beneficial to introduce an intermediate layer, a first principles-based reactive force field (ReaxFF) approach,[18,19] in which interatomic interaction adapts dynamically to the local environment to describe chemical reactions. The ReaxFF is trained by performing thousands of small QM calculations.

The metascalable simulation framework is enabling the study of a number of exciting problems, in particular, how atomistic processes determine material properties. Examples include the mechanical properties of nanocomposite materials and nanoindentation on them,[20] oxidation of nanoenergetic materials,[21] hypervelocity impact damage,[22] and fracture.[23,24] We also study both colloidal[25] and epitaxial[26] quantum dots, and their interface with biological systems. It is noteworthy that experimentalists can now observe these phenomena at the same resolution as our simulations. For example, experimentalists perform nano-shock experiments using focused laser beams[27] and nano-fracture measurements using atomic force microscopy.[28] This lecture note focuses on one application related to nano-mechano-chemistry, i.e., enhanced reaction mechanisms in nanostructured energetic materials.

The lecture note is organized as follows. Section 2 describes our metascalable computing framework for large spatiotemporal-scale simulations of chemical reactions based on spatiotemporal data-locality principles. Results of nano-mechano-chemistry simulations are given in section 3, and section 4 contains conclusions.

2 A Metascalable Dwarf

2.1 Embedded Divide-and-Conquer (EDC) Algorithmic Framework

In the embedded divide-and-conquer (EDC) algorithms, the physical system is divided into spatially localized computational cells.[2] These cells are embedded in a global field that is computed efficiently with tree-based algorithms (Fig. 1).

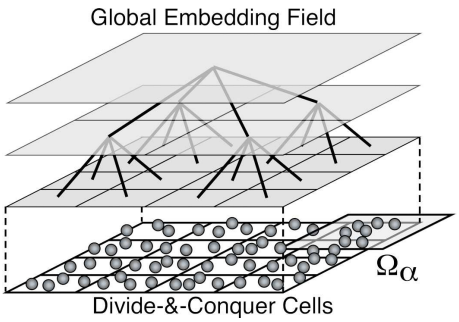


Figure 1. Schematic of embedded divide-and-conquer (EDC) algorithms. The physical space is subdivided into spatially localized cells, with local atoms constituting subproblems, which are embedded in a global field solved with tree-based algorithms.

Within the EDC framework, we have designed a number of $O(N)$ algorithms (N is the number of atoms). For example, we have designed a space-time multiresolution MD (MRMD) algorithm to reduce the $O(N^2)$ complexity of the N-body problem to $O(N)$.[13] MD simulation follows the trajectories of N point atoms by numerically integrating coupled ordinary differential equations. The hardest computation in MD simulation is the evaluation of the long-range electrostatic potential at N atomic positions. Since each evaluation involves contributions from $N-1$ sources, direct summation requires $O(N^2)$ operations. The MRMD algorithm uses the octree-based fast multipole method (FMM)[29,30] to reduce the computational complexity to $O(N)$ based on spatial locality. We also use multiresolution in time, where temporal locality is utilized by computing forces from further atoms with less frequency.[31]

We have also designed a fast ReaxFF (F-ReaxFF) algorithm to solve the $O(N^3)$ variable N-charge problem in chemically reactive MD in $O(N)$ time.[19] To describe chemical bond breakage and formation, the ReaxFF potential energy is a function of the positions of atomic pairs, triplets and quadruplets as well as the chemical bond orders of all constituent atomic pairs.[18] To describe charge transfer, ReaxFF uses a charge-equilibration scheme, in which atomic charges are determined at every MD step to minimize the electrostatic energy with the charge-neutrality constraint. This variable N-charge problem amounts to solving a dense linear system of equations, which requires $O(N^3)$ operations. The F-ReaxFF algorithm uses the FMM to perform the matrix-vector multiplications with $O(N)$ operations. It further utilizes the temporal locality of the solutions to reduce the amortized computational cost averaged over simulation steps to $O(N)$. To further speed up the solution, we use a multilevel preconditioned conjugate gradient (MPCG) method.[21,32] This method splits the

Coulomb interaction matrix into far-field and near-field matrices and uses the sparse near-field matrix as a preconditioner. The extensive use of the sparse preconditioner enhances the data locality, thereby increasing the parallel efficiency.

To approximately solve the exponentially complex quantum N-body problem in $O(N)$ time,[33,34] we use an EDC density functional theory (EDC-DFT) algorithm.[17,35] The DFT reduces the exponential complexity to $O(N^3)$, by solving N_{el} one-electron problems self-consistently instead of one N_{el}-electron problem (the number of electrons, N_{el}, is on the order of N). The DFT problem can be formulated as a minimization of an energy functional with respect to electronic wave functions. In the EDC-DFT algorithm, the physical space is a union of overlapping domains, $\Omega = \sum_\alpha \Omega_\alpha$ (Fig. 1), and physical properties are computed as linear combinations of domain properties that in turn are computed from local electronic wave functions. For example, the electronic density $\rho(\mathbf{r})$ is calculated as $\rho(\mathbf{r}) = \sum_\alpha p^\alpha(\mathbf{r}) \sum_n f(\epsilon_n^\alpha)|\psi_n^\alpha(\mathbf{r})|^2$, where the support function $p^\alpha(\mathbf{r})$ vanishes outside domain $\Omega_\alpha$and satisfies the sum rule, $\sum_\alpha p^\alpha(\mathbf{r}) = 1$, and $f(\epsilon_n^\alpha)$ is the Fermi distribution function corresponding to the energy ϵ_n^α of the n-th electronic wave function (or Kohn-Sham orbital) $\psi_n^\alpha(\mathbf{r})$ in Ω_α . For DFT calculation within each domain, we use a real-space approach based on high-order finite differencing,[36] where iterative solutions are accelerated using the multigrid preconditioning.[37] The multigrid is augmented with high-resolution grids that are adaptively generated near the atoms to accurately operate atomic pseudopotentials.[17] The numerical core of EDC-DFT thus represents a high-order stencil computation.[38,39]

2.2 Space-Time-Ensemble Parallelism (STEP) for Predicting Long-Time Dynamics

A challenging problem is to predict long-time dynamics because of the sequential bottleneck of time.[4,5] Due to temporal locality, however, the system stays near local minimum-energy configurations most of the time, except for rare transitions between them. In such cases, the transition state theory (TST) allows the reformulation of the sequential long-time dynamics as computationally more efficient parallel search for low activation-barrier transition events.[10,40] We also introduce a discrete abstraction based on graph data structures, so that combinatorial techniques can be used for the search.[40] We have developed a directionally heated nudged elastic band (DH-NEB) method,[9] in which a NEB consisting of a sequence of S states,[41] $\mathbf{R}_s \in \Re^{3N} (s = 0, ..., S-1, \Re$ is the set of real numbers, and N is the number of atoms), at different temperatures searches for transition events (Fig. 2(a)):

$$\mathbf{M}\frac{d^2}{dt^2}\mathbf{R}_s = \mathbf{F}_s - \mathbf{M}\gamma_s\frac{d}{dt}\mathbf{R}_s \tag{1}$$

where $\mathbf{M} \in \Re^{\mathbf{3N}\times\mathbf{3N}}$ is the diagonal mass matrix and γ_s is a friction coefficient. Here, the forces are defined as

$$\mathbf{F}_s = \begin{cases} -\frac{\partial V}{\partial \mathbf{R}_s}|_\perp + \mathbf{F}_s^{spr}|_\parallel \\ -\frac{\partial V}{\partial \mathbf{R}_s} \end{cases} \tag{2}$$

where $V(\mathbf{R})$ is the interatomic potential energy, $\mathbf{F}_s^{spr}$ are spring forces that keep the states equidistance, and $\perp$ and $\parallel$ denote respectively the projections of a $3N$-element vector perpendicular and parallel to the tangential vector connecting the consecutive states.

We use an ensemble consisting of B bands to perform long-time simulation—molecular kinetics (MK) simulation–in the framework of kinetic Monte Carlo simulation.[9] Here, our space-time-ensemble parallel (STEP) approach combines a hierarchy of concurrency, i.e., the number of processors is

$$P = BSD : \tag{3}$$

(1) spatial decomposition within each state (D is the number of spatial subsystems, see section 2.3); (2) temporal parallelism across S states within each band; and (3) ensemble parallelism over B bands (Fig. 2(b)).

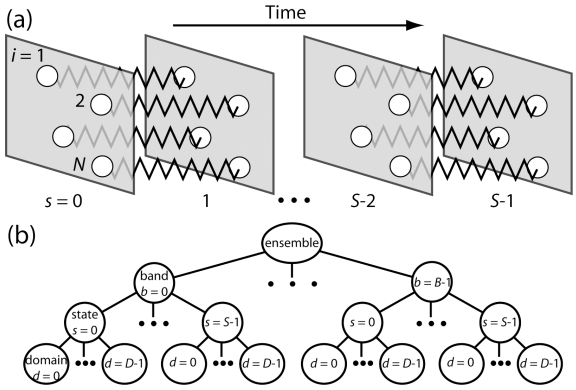


Figure 2. Schematic of the space-time-ensemble parallel (STEP) approach. (a) A nudged elastic band consists of a sequence of S states (gray parallelograms), $\mathbf{R}_s$ (s = 0,...,$S-1$), where each state consists of N atoms (white spheres), i = 1,...,N. Corresponding atoms in consecutive states interact via harmonic forces represented by wavy lines. (b) Tree-structured processor organization in the STEP approach. An ensemble consists of B bands, each consisting of S states; each state in turn contains D spatial domains.

2.3 Tunable Hierarchical Cellular Decomposition (HCD) for Algorithm-Hardware Mapping

To map the $O(N)$ EDC-STEP algorithms onto parallel computers, we have developed a tunable hierarchical cellular decomposition (HCD) framework.

Massively distributed scalability via message passing—Superscalability: Our parallelization in space is based on spatial decomposition, in which each spatial subsystem is assigned to a compute node in a parallel computer. For large granularity (the number of atoms per spatial subsystem, $N/D > 10^2$), simple spatial decomposition (i.e., each node is responsible for the computation of the forces on the atoms within its subsystem) suffices, whereas for finer granularity ($N/D \sim 1$), neutral-territory[5,42] or other hybrid decomposition schemes[4,43–45] can be incorporated into the framework. Our parallelization framework also includes load-balancing capability. For irregular data structures, the number of atoms assigned to each processor varies significantly, and this load imbalance degrades the parallel efficiency. Load balancing can be stated as an optimization

problem.[46–48] We minimize the load-imbalance cost as well the size and the number of messages. Our topology-preserving spatial decomposition allows message passing to be performed in a structured way in only 6 steps, so that the number of messages is minimized. To minimize the load imbalance cost and the size of messages, we have developed a computational-space decomposition scheme.[49] The main idea is that the computational space shrinks in a region with high workload density, so that the workload is uniformly distributed. The sum of load-imbalance and communication costs is minimized as a functional of the computational space using simulated annealing. We have found that wavelets allow compact representation of curved partition boundaries and thus speed up the optimization procedure.[50]

Multicore scalability via multithreading—Nanoscalablity: In addition to the massive inter-node scalability, "there is plenty of room at the bottom," as Richard Feynman noted. At the finest level, EDC algorithms consist of a large number of computational cells (Fig. 1), such as linked-list cells in MD[13] and domains in EDC-DFT,[17] which are readily amenable to parallelization. On a multicore compute node, a block of cells is assigned to each thread for intra-node parallelization. Our EDC algorithms are thus implemented as hybrid message passing + multithreading programs. Here, we use the POSIX thread standard, which is supported across broad architectures and operating systems. In addition, our framework[2] includes the optimization of data and computation layouts,[51,52] in which the computational cells are traversed along various spacefilling curves[53] (e.g. Hilbert or Morton curve). To achieve high efficiency, special care must be taken also to make the multithreading free of critical sections. For example, we have designed a critical section-free algorithm to make all interatomic force computations in MRMD independent by reorganization of summation of atomic pair and triplet summations.[12] Our multithreading is based on a master/worker model, in which a master thread coordinates worker threads that actually perform force computations. We use POSIX semaphores to signal between the master and worker threads to avoid the overhead of thread creation and joining in each MD step. There are two check points at each MD time step, where all worker threads wait a signal from the master thread: (1) before the two-body force calculation loop, which also constructs the neighbor-lists, after atomic coordinates are updated; and (2) before three-body force calculation, after having all atoms complete neighbor-list construction. We have also combined multithreading with single-instruction multiple-data (SIMD) parallelism based on various code transformations.[39] Our SIMD transformations include translocated statement fusion, vector composition via shuffle, and vectorized data layout reordering (e.g. matrix transpose), which are combined with traditional optimization techniques such as loop unrolling.

Long-time scalability via space-time-ensemble parallelism (STEP)—Eon-scalability: With the spatial decomposition, the computational cost scales as N/D, while communication scales in proportion to $(N/D)^{2/3}$.[13] For long-range interatomic potentials used in MD simulations, tree-based algorithms such as the fast multipole method (FMM)[29,30] incur an $O(\log D)$ overhead, which is negligible for coarse grained $(N/D \gg D)$ applications.[30] The communication cost of the temporal decomposition is $O(N/D)$ for copying nearest-neighbor images along the temporal axis, but the prefactor is negligible compared with the computation. Ensemble decomposition duplicates the band calculation, each involving SD processors, B times on $P = BSD$ processors. It involves $O((N/D)\log(BS))$ overhead to multicast the new initial state among the

processors assigned the same spatial domain, i.e., those with the same p mod D.[9] Here, $p = bSD + sD + d$ is the sequential processor ID, where processor p is assigned the d-th spatial subsystem of the s-th state in the b-th band. The multicast cost at the beginning of each molecular-kinetics (MK) simulation step is greatly amortized over $10^3 - 10^4$ MD steps performed for the DH-NEB method per MK iteration.[9]

Intelligent tuning: The hierarchy of computational cells provides an efficient mechanism for performance optimization as well we make both the layout and size of the cells as tunable parameters that are optimized on each computing platform.[2] Our EDC-STEP algorithms are implemented as hybrid message-passing + multithreading programs in the tunable HCD framework, in which the numbers of message passing interface (MPI) processes and POSIX threads are also tunable parameters. The HCD framework thus maximally exposes data locality and concurrency. We are currently collaborating with compiler and artificial intelligence (AI) research groups to use: (1) knowledge-representation techniques for expressing the exposed concurrency; and (2) machine-learning techniques for optimally mapping the expressed concurrency to hardware.[54]

2.4 Scalability Tests

The scalability of our EDC-STEP-HCD applications has been tested on various high-end computing platforms including 212,992 IBM BlueGene/L processors at the Lawrence Livermore National Laboratory and 131,072 IBM BlueGene/P processors at the Argonne National Laboratory.

Inter-node (message-passing) spatial scalability: Figure 3 shows the execution and communication times of the MRMD, F-ReaxFF and EDC-DFT algorithms as a function of the number of processors P on the IBM BlueGene/L and P. Figure 3(a) shows the execution time of the MRMD algorithm for silica material as a function of P. We scale the problem size linearly with the number of processors, so that the number of atoms N = 2,044,416P. In the MRMD algorithm, the interatomic potential energy is split into the long- and short-range contributions, and the long-range contribution is computed every 10 MD time steps. The execution time increases only slightly as a function of P on both BlueGene/L and P, and this signifies an excellent parallel efficiency. We define the speed of an MD program as a product of the total number of atoms and time steps executed per second. The isogranular speedup is the ratio between the speed of P processors and that of one processor. The weak-scaling parallel efficiency is the speedup divided by P, and it is 0.975 on 131,072 BlueGene/P processors. The measured weak-scaling parallel efficiency on 212,992 BlueGene/L processors is 0.985 based on the speedup over 4,096 processors. Figure 3(a) also shows that the algorithm involves very small communication time. Figure 3(b) shows the execution time of the F-ReaxFF MD algorithm for RDX material as a function of P, where the number of atoms is N = 16,128P. The computation time includes 3 conjugate gradient (CG) iterations to solve the electronegativity equalization problem for determining atomic charges at each MD time step. On 212,992 BlueGene/L processors, the isogranular parallel efficiency of the F-ReaxFF algorithm is 0.996. Figure 3(c) shows the performance of the EDC-DFT based MD algorithm for 180P atom alumina systems. The execution time includes 3 self-consistent (SC) iterations to determine the electronic wave functions and the Kohn-Sham potential, with 3 CG iterations per SC cycle to refine each wave function iteratively. On 212,992 BlueGene/L processors, the isogranular parallel efficiency of the

EDC-DFT algorithm is 0.998 (based on the speedup over 4,096 processors). Our largest benchmark tests include 217,722,126,336-atom MRMD, 1,717,567,488-atom F-ReaxFF, and 19,169,280-atom (1,683,216,138,240 electronic degrees-of-freedom) EDC-DFT calculations on 212,992 BlueGene/L processors.

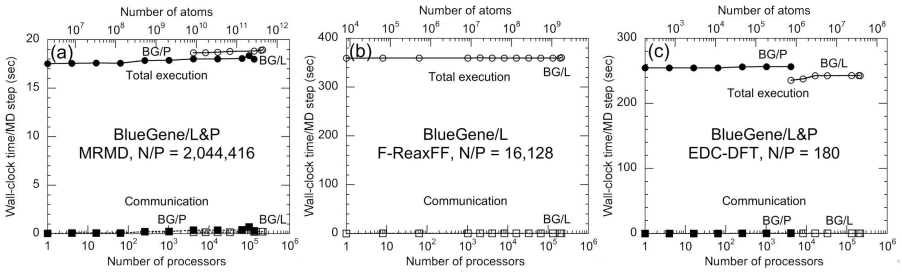


Figure 3. Total execution (circles) and communication (squares) times per MD time step as a function of the number of processors P of BlueGene/L (open symbols) and BlueGene/P (solid symbols) for three MD simulation algorithms: (a) MRMD for 2,044,416P atom silica systems; (b) F-ReaxFF MD for 16,128P atom RDX systems; and (c) EDC-DFT MD for 180P atom alumina systems.

Intra-node (multithreading) spatial scalability: We have tested the multithreading scalability of MRMD on a dual Intel Xeon quadcore platform. Figure 4 shows the speedup of the multithreaded code over the single-thread counterpart as a function of the number of worker threads. In addition to the speedup of the total program, Fig. 4 also shows the speedups of the code segments for two-body and three-body force calculations separately. We see that the code scales quite well up to 8 threads on the 8-core platform. We define the multithreading efficiency as the speedup divided by the number of threads. The efficiency of two-body force calculation is 0.927, while that for three-body force calculation is 0.436, for 8 threads. The low efficiency of the three-body force calculation may be due to the redundant computations introduced to eliminate critical sections. Nevertheless, the efficiency of the total program is rather high (0.811), since the fraction of the three-body calculation is about one third of the two-body force calculation. This result shows that the semaphore-based signaling between master and worker threads is highly effective. In a test calculation for a 12,228-atom silica system, the running time is 13.6 milliseconds per MD time step.

Time/ensemble scalability Scalability of the STEP-MRMD algorithm (note that the STEP approach can be combined with any of the MRMD, F-ReaxFF and EDC-DFT algorithms to compute interatomic forces) is tested on a cluster of dual-core, dual-processor AMD Opteron (at clock frequency 2 GHz) nodes with Myrinet interconnect. We define the speed of a program as a product of the total number of atoms and MK simulation steps executed per second. The speedup is the ratio between the speed of P processors and that of one processor. The parallel efficiency is the speedup divided by P. We first test the scalability of temporal decomposition, where we fix the number of bands B = 1 and the number of domains per state D = 1. We vary the number of states per band S = 4 to 1024. Here, the simulated system is amorphous SiO_2 consisting of N = 192 atoms, and we perform 600 MD steps per MK simulation step. The test uses all four cores per node. Figure 5(a) shows

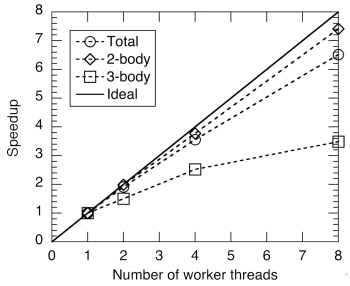


Figure 4. Speedup of the multithreaded MRMD algorithm over a single-threaded counterpart for the total program (circles), the two-body force calculation (diamonds), and three-body force calculation (squares). The solid line shows the ideal speedup.

the speedup of the STEP-MRMD program (we normalize the speedup on 4 processors as 4). The measured speedup on 1,024 processors is 980.2, and thus the parallel efficiency is 0.957. Next, we test the scalability of ensemble decomposition, where we fix the number of states per band S = 4 and the number of spatial domains per state D = 1. The number of bands per ensemble is varied from B = 1 to 256. The simulated system is amorphous SiO_2 consisting of N = 192 atoms. Although multiple events are generated independently by different processor groups, the parallel algorithm involves sequential bottlenecks such as the selection of an event that occurs, and accordingly the parallel efficiency does degrade for a larger number of processors. Figure 5(b) shows the speedup of the STEP-MRMD program on the Opteron cluster as a function of the number of processors (normalized to be 4 on 4 processors). On 1,024 processors, the measured speedup is 989.2, and thus the parallel efficiency of ensemble decomposition is 0.966, which is slightly higher than that of temporal decomposition on the same number of processors.

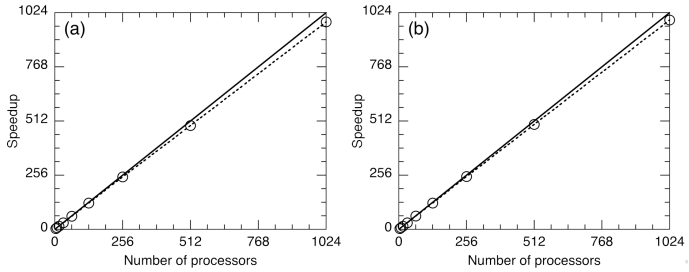


Figure 5. (a) Speedup of temporal decomposition in the STEP-MRMD algorithm (normalized so that the speedup is 4 for P = 4) as a function of the number of processors P (P = 4-1024) for a 192-atom amorphous SiO2 system on dual-core, dual-processor AMD Opteron nodes, where we fix $B = D$ = 1. The circles are measured speedups, whereas the solid line denotes the perfect speedup. (b) Speedup of ensemble decomposition in the STEP-MRMD algorithm as a function of the number of processors P (= 4,...,1024) for silica material (N = 192 atoms). Here, we fix the number of states per band S = 4 and the number of spatial domains per state D = 1, while the number of bands is varied from B = 1 to 256.

3 Nano-Mechano-Chemistry Simulations

Recent advances in the integration of nanowires and nanoparticles of energetic materials into semiconducting electronic structures have opened up the possibility of "nanoenergetics-on-a-chip (NOC)" technology, which has a wide range of potential applications such as micropropulsion in space and nano-airbags to drive nanofluidics.[55] Most widely used energetic materials for device integration are thermites, which are composites of metals and oxides. These materials have enormous energy release associated with the highly exothermic reduction/oxidation (redox) reactions to form more stable oxides. For example, arrays of Fe_2O_3 and CuO nanowires embedded in an Al matrix have been deposited on solid surfaces.[56] Another example of thermite nanostructures is self-assembly of an ordered array of Al and Fe_2O_3 nanoparticles.[57]

The integration of nanoenergetic materials into electronic circuits requires fundamental understanding and precise control of reaction rates and initiation time. The reactivity of nanoenergetic materials is known to differ drastically from their micron-scale counterparts. For example, experimental studies on the combustion of nanothermites, such as Al/Fe_2O_3, have shown that flame propagation speeds approach km/s when the size of Al nanoparticles is reduced to below 100 nm, in contrast to cm/s for traditional thermites.[58] Another example is the two-stage reaction of Al/CuO-nanowire thermite, in which the first reaction takes place at 500 °C followed by the second reaction at 660 °C (i.e., Al melting temperature).[56]

Such peculiar reactive behaviors of nanothermites cannot be explained by conventional mechanisms based on mass diffusion of reactants, and thus various alternative mechanisms have been proposed. An example is a mechano-chemical mechanism that explains the fast flame propagation based on dispersion of the molten metal core of each nanoparticle and spallation of the oxide shell covering the metal core.[59] Another mechanism is accelerated mass transport of both oxygen and metal atoms due to the large pressure gradient between the metal core and the oxide shell of each metal nanoparticle.[21,60] In addition, defect-mediated giant diffusivity is important for fast reactions at the nanometer scale.[24,61,62]

The above mechanisms are by no means exhaustive, and some unexpected ones could operate in NOCs. It is therefore desirable to study the reaction of nanoenergetic materials by first-principles simulations. However, this poses an enormous theoretical challenge, where quantum-mechanical accuracy to describe chemical reactions must be combined with large spatial scales to capture nanostructural effects. Recent developments in scalable reactive MD simulations as described in the previous section have set the stage for such large first-principles MD simulations.

We have performed embedded divide-and-conquer (EDC) density functional theory (DFT) based MD simulations to study the thermite reaction at an Al/Fe_2O_3 interface (Fig. 6).[63] The results reveal a concerted metal-oxygen flip mechanism that significantly enhances the rate of redox reactions. This mechanism leads to two-stage reactions—rapid initial reaction due to collective metal-oxygen flips followed by slower reaction based on uncorrelated diffusive motions, which may explain recent experimental observation in thermite nanowire arrays.[56]

Here, we simulate a stack of Al and Fe_2O_3 layers involving 1,152 (144 Fe_2O_3 + 432 Al) atoms with periodic boundary conditions. The hematite (Fe_2O_3) crystal, cut along (0001) planes to expose Fe planes, is placed in the supercell with the (0001) direction parallel to the z direction (Fig. 6(a)). The Fe planes of the hematite are attached to (111)

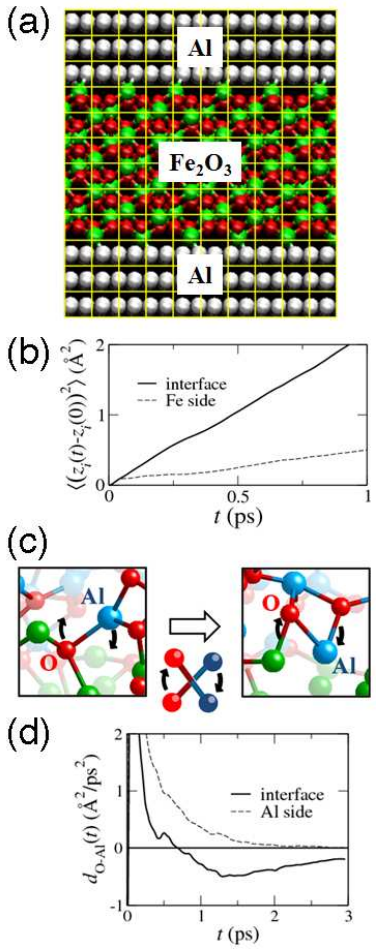


Figure 6. (a) Atomic configuration of Al/Fe_2O_3 interface. The green, red and grey spheres show the positions of Fe, O and Al atoms, respectively. Yellow meshes show the non-overlapping cores used by the EDC-DFT method. (b) Enhanced diffusion at the metal-oxide interface. Mean square displacements of O atoms along the z direction are plotted as a function of time. The solid and dashed curves are for O atoms in the interfacial and Fe-side regions, respectively. (c) Concerted metal-oxygen flip at the Al/Fe_2O_3 interface. (d) Negative correlation associated with concerted Al and O motions at the interface. Correlation functions between displacements of O and Al atoms along the z direction are shown as a function of time. The solid and dashed curves are obtained in the interfacial and Al-side regions.

planes of the face-centered cubic Al crystal at the two interfaces. Simulation results show enhanced mass diffusivity at the metal/oxide interface (Fig. 6(b)). To understand the mechanism of the enhanced diffusivity at the interface, we have examined the time evolution of the atomic configuration in the interfacial region and found a concerted metal-oxygen flip mechanism (Fig. 6(c)). That is, O atoms switch their positions with neighboring Al atoms while diffusing in the z direction. Careful bond-overlap population analysis shows that the switching motion between O and Al atoms at the interface is triggered by the change of

chemical bonding associated with these atoms. To quantify the collective switching motion between O and Al atoms, we calculate the correlation function between the displacements of atoms along the z direction. The results in Fig. 6(d) (solid curve) reveal negative correlation for $t > 0.5$ ps, which reflects the collective switching motion between O and Al atoms at the interface as shown in Fig. 6(c). Such negative correlation does not exist on the Al side (the dashed curve in Fig. 6(d)), indicating independent diffusive motions of Al and O atoms.

Reactivity of nanoenergetic materials is often enhanced drastically from their micron-scale counterparts, which cannot be explained by conventional mechanisms based on mass diffusion of reactants. We have studied atomistic mechanisms of oxidation of an aluminum nanoparticle under extreme environment using multimillion atom reactive (ReaxFF) MD simulations, where the aluminum nanoparticle is coated with crystalline alumina shell and is ignited by heating the aluminum core to high temperatures, as is done in recent laser flash-heating experiments (Fig. 7).[27] The metallic aluminum and ceramic alumina are modeled by embedded atom model and many-body ionic-covalent potential form, respectively, which are interpolated with a bond-order based scheme validated quantum mechanically.

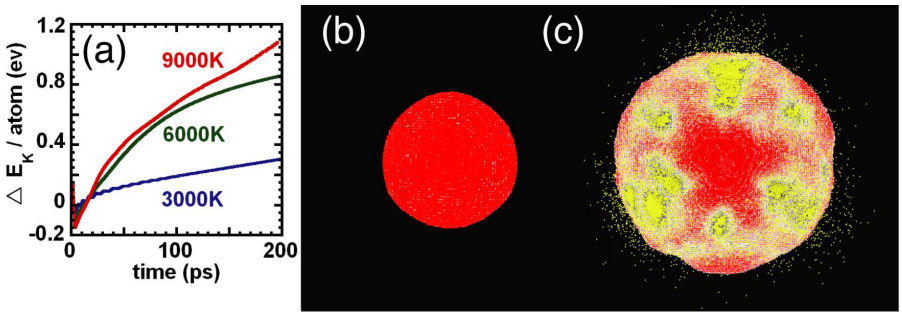


Figure 7. (a) Time variation of kinetic energy per aluminum atom during explosion with different initial temperatures T=3,000K (blue), 6,000K (green), and 9,000K (red), respectively. (b) Snapshot of nanoparticle at 100 ps (T=3,000K). (c) Snapshot of nanoparticle at 100 ps (T=9,000K). Core Al atoms (yellow) jet out through holes on the nanoparticle shell (red).

Simulation results reveal a transition of the reaction mechanism from thermodynamic to mechano-chemical regime, resulting in faster oxidation reaction of the aluminum nanoparticle, at elevated temperatures (Fig. 7(a)). The breakdown of the shell and the change of shell's morphology and composition during oxidation are found to play an important role for the transition. Specifically, we have identified three major changes of the shell, which are related to three mechanisms of atom migration: Diffusion (Fig. 7(b)), ballistic transport followed by diffusion, and ballistic transport followed by coalescing of atoms into few-atom clusters (Fig. 7(c)).

Mechanical stimuli in energetic materials initiate chemical reactions at shock fronts prior to detonation. Shock sensitivity measurements provide widely varying results, and quantum mechanical calculations are unable to handle systems large enough to describe shock structure. Recent developments in ReaxFF-MD combined with advances in parallel

computing have paved the way to accurately simulate reaction pathways along with the structure of shock fronts. Our multimillion-atom ReaxFF-MD simulations of 1,3,5-trinitro-1,3,5-triazine (RDX) (Figs. 8(a) and (b)) reveal that detonation is preceded by a transition from a diffuse shock front with well ordered molecular dipoles behind it to a disordered dipole distribution behind a sharp front.[3]

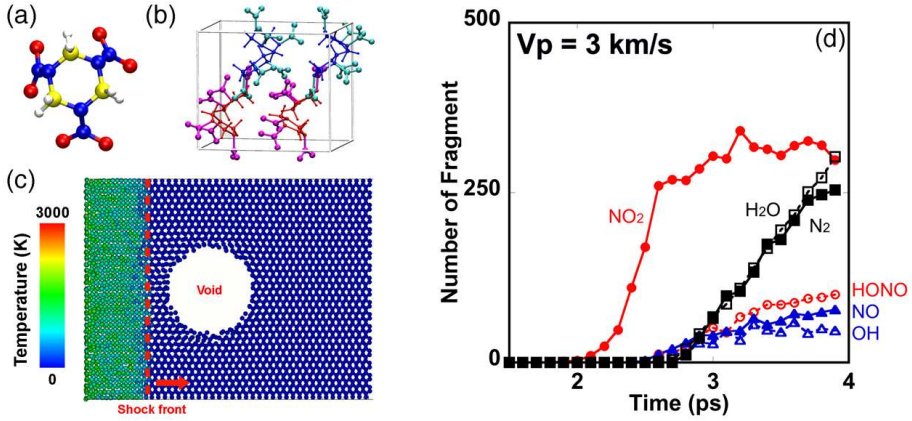


Figure 8. (a) An RDX molecule with carbon (yellow), hydrogen (white), oxygen (red), and nitrogen (blue) atoms. (b) The unit cell of an RDX crystal contains 8 RDX molecules, which are colored blue and red depending on whether the NO_2 groups faces away from (group1) or faces towards (group2) the shock plane. (c) Distribution of molecular vibrational temperature around the void at a particle velocity of 3 km/s. A red dotted-line represents the position of shock front. (d) Number of molecular fragments near the void surface. As the void collapses, two distinct reaction regimes are observed. From the arrival of the shock wave until the void closure ($\sim$ 2.6 ps), a rapid production of NO_2 is observed. Shortly after that, when molecules strike the downstream wall (2.6 – 3.9 ps), various chemical products such as N_2, H_2O and HONO are produced.

Nanofluidics of chemically reactive species has enormous technological potential and computational challenge arising from coupling quantum-mechanical accuracy with large-scale fluid phenomena. We have performed multimillion-atom ReaxFF-MD simulation of shock initiation of an RDX crystal with a nanometer-scale void (Fig. 8(c)).[64] The simulation reveals the formation of a nanojet that focuses into a narrow beam at the void. This, combined with the excitation of vibrational modes through enhanced intermolecular collisions by the free volume of the void, catalyzes chemical reactions that do not occur otherwise (Fig. 8(d)). We also observe a pinning-depinning transition of the shock wave front at the void at increased particle velocity and the resulting localization-delocalization transition of the vibrational energy. More recently, we have simulated nanoindentation of the (100) crystal surface of RDX by a diamond indenter.[65] Nanoindentation causes significant heating of the RDX substrate in the proximity of the indenter, resulting in the release of molecular fragments and subsequent ”walking” motion of these molecules on the indenter surfaces.

4 Conclusions

In summary, we have developed high-end reactive atomistic simulation programs to encompass large spatiotemporal scales with common algorithmic and computational frameworks based on spatiotemporal data-locality principles. In fact, our "metascalable dwarf" extends far beyond atomistic simulations: Diverse applications, which encompass all of the original seven dwarfs, can be reduced by common techniques of embedding and divide-and-conquer to a highly scalable form. According to the scalability tests presented in this lecture note, they are likely to scale on future architectures beyond petaflops. The simulation algorithms are already enabling million-to-billion atom simulations of mechano-chemical processes, which have applications in broad areas such as energy and environment.

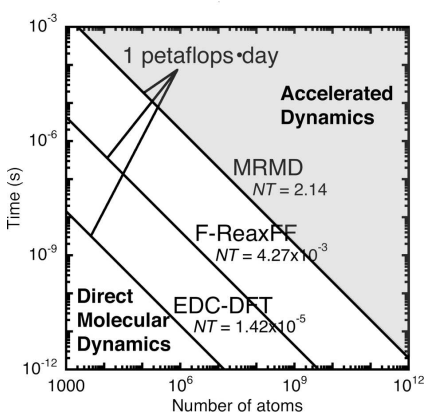


Figure 9. Spatiotemporal scales NT accessible by direct molecular-dynamics (white background) and approximate accelerated-dynamics (gray) simulations with a petaflops•day of computing. The lines are the NT achieved per petaflopsday of computing for MD (MRMD), chemically reactive MD (F-ReaxFF), and quantum-mechanical MD (EDC-DFT) simulations, respectively.

A critical issue, however, is the time scale studied by MD simulations. We define the spatiotemporal scale, NT, of an MD simulation as the product of the number of atoms N and the simulated time span T. On petaflops computers, direct MD simulations can be performed for NT = 1–10 atomseconds (i.e. multibillion-atom simulation for several nanoseconds or multimillion-atom simulation for several microseconds). More specifically, a day of computing on a sustained petaflops computer (i.e. one petaflops•day of computing) achieves NT = 2.14 (e.g. 1 million atoms for 2.14 microseconds) (Fig. 9), according to the benchmark test in section 2 (i.e., extrapolated from the measured MRMD performance on the BlueGene/L, which is rated as 0.478 petaflops according to the Linpack benchmark).[12] Accelerated-dynamics simulations[10] such as STEP molecular-kinetics simulations[9] will push the spatiotemporal envelope beyond NT = 10, but they need to be fully validated against direct MD simulations at NT = 1–10. Such large spatiotemporal-scale atomistic simulations are expected to advance scientific knowledge. This work was supported by NSF-ITR/PetaApps/EMT, DOE-SciDAC/BES, ARO-MURI, DTRA, and Chevron-CiSoft.

We thank Richard Clark, Liu Peng, Richard Seymour, and Lin H. Yang for fruitful collaborations.

References

1. S. Emmott and S. Rison, Towards 2020 Science (Microsoft Research, Cambridge, UK, 2006).
2. A. Nakano et al., Int'l J High Performance Comput Appl **22**, 113 (2008).
3. K. Nomura et al., Phys Rev Lett **99**, 148303 (2007).
4. J. C. Phillips et al., in Proc of Supercomputing (SC02) (ACM/IEEE, 2002).
5. D. E. Shaw et al., ACM SIGARCH Computer Architecture News **35**, 1 (2007).
6. K. Asanovic et al., The Landscape of Parallel Computing Research: A View from Berkeley (University of California, Berkeley, 2006).
7. J. Dongarra et al., CTWatch Quarterly **3**, 11 (2007).
8. A. Nakano et al., Comput Mater Sci **38**, 642 (2007).
9. A. Nakano, Comput Phys Commun **178**, 280 (2008).
10. A. F. Voter, F. Montalenti, and T. C. Germann, Annual Rev Mater Res **32**, 321 (2002).
11. F. Shimojo et al., J Phys: Condens Matter **20**, 294204 (2008).
12. K. Nomura et al., in Proc of International Parallel and Distributed Processing Symposium (IPDPS) (IEEE, 2009).
13. A. Nakano et al., in Proc of Supercomputing (SC01) (ACM/IEEE, 2001).
14. J. N. Glosli et al., in Proc of Supercomputing (SC07) (ACM/IEEE, 2007).
15. J. Nieplocha, R. J. Harrison, and R. J. Littlefield, in Proc of Supercomputing (SC94) (ACM/IEEE, 1994).
16. F. Gygi et al., in Proc of Supercomputing (SC05) (ACM/IEEE, 2005).
17. F. Shimojo et al., Phys Rev B **77**, 085103 (2008).
18. A. C. T. van Duin et al., J Phys Chem A **105**, 9396 (2001).
19. K. Nomura et al., Comput Phys Commun **178**, 73 (2008).
20. I. Szlufarska, A. Nakano, and P. Vashishta, Science **309**, 911 (2005).
21. T. J. Campbell, et al., Phys Rev Lett **82**, 4866 (1999).
22. P. S. Branicio et al., Phys Rev Lett **96**, 065502 (2006).
23. Z. Lu et al., Phys Rev Lett **95**, 135501 (2005).
24. Y. Chen et al., Phys Rev Lett **99**, 155506 (2007).
25. S. Kodiyalam et al., Phys Rev Lett **93**, 203401 (2004).
26. E. Lidorikis et al., Phys Rev Lett **87**, 086104 (2001).
27. Y. Q. Yang et al., Appl Phys Lett **85**, 1493 (2004).
28. F. Celarie et al., Phys Rev Lett **90**, 075504 (2003).
29. L. Greengard and V. Rokhlin, J Comput Phys **73**, 325 (1987).
30. S. Ogata et al., Comput Phys Commun **153**, 445 (2003).
31. A. Nakano, R. K. Kalia, and P. Vashishta, Comput Phys Commun **83**, 197 (1994).
32. A. Nakano, Comput Phys Commun **104**, 59 (1997).
33. S. Goedecker, Rev Mod Phys **71**, 1085 (1999).
34. D. R. Bowler et al., J Phys: Condens Matter **20**, 290301 (2008).
35. W. T. Yang, Phys Rev Lett **66**, 1438 (1991).
36. J. R. Chelikowsky et al., Physica Status Solidi B **217**, 173 (2000).
37. J.-L. Fattebert and J. Bernholc, Phys Rev B **62**, 1713 (2000).

38. K. Datta et al., in Proc of Supercomputing (SC08) (ACM/IEEE, 2008).
39. L. Peng et al., in Proc of International Parallel and Distributed Processing Symposium (IPDPS) (IEEE, 2009).
40. A. Nakano, Comput Phys Commun **176**, 292 (2007).
41. G. Henkelman and H. Jonsson, J Chem Phys **113**, 9978 (2000).
42. D. E. Shaw, J Comput Chem **26**, 1318 (2005).
43. S. J. Plimpton, J Comput Phys **117**, 1 (1995).
44. B. G. Fitch et al., Lecture Notes in Computer Science **3992**, 846 (2006).
45. M. Snir, Theor Comput Sys **37**, 295 (2004).
46. R. D. Williams, Concurrency: Practice and Experience **3**, 457 (1991).
47. K. D. Devine et al., Appl Numer Math **52**, 133 (2005).
48. U. V. Catalyurek et al., in Proc of International Parallel and Distributed Processing Symposium (IPDPS) (IEEE, 2007).
49. A. Nakano and T. J. Campbell, Parallel Comput **23**, 1461 (1997).
50. A. Nakano, Concurrency: Practice and Experience **11**, 343 (1999).
51. J. Mellor-Crummey, D. Whalley, and K. Kennedy, Int'l J Parallel Prog **29**, 217 (2001).
52. M. M. Strout and P. D. Hovland, in Proc of the Workshop on Memory System Performance (ACM, 2004).
53. B. Moon et al., IEEE Trans Knowledge Data Eng **13**, 124 (2001).
54. B. Bansal et al., in Proc of the Next Generation Software Workshop, International Parallel and Distributed Processing Symposium (IPDPS) (IEEE, 2007).
55. C. Rossi et al., J Microelectromech Sys **16**, 919 (2007).
56. K. Zhang et al., Appl Phys Lett **91**, 113117 (2007).
57. S. H. Kim and M. R. Zachariah, Adv Mater **16**, 1821 (2004).
58. K. B. Plantier, M. L. Pantoya, and A. E. Gash, Combustion and Flame **140**, 299 (2005).
59. V. I. Levitas et al., Appl Phys Lett **89**, 071909 (2006).
60. A. Rai et al., Combustion Theory and Modelling **10**, 843 (2006).
61. N. N. Thadhani, J Appl Phys **76**, 2129 (1994).
62. M. Legros et al., Science **319**, 1646 (2008).
63. F. Shimojo et al., Phys Rev E **77**, 066103 (2008).
64. K. Nomura et al., Appl Phys Lett **91**, 183109 (2007).
65. Y. Chen et al., Appl Phys Lett **93**, 171908 (2008).

Soft Matter, Fundamentals and Coarse Graining Strategies

Christine Peter and Kurt Kremer

Max Planck Institute for Polymer Research
Ackermannweg 10, 55128 Mainz, Germany
E-mail: {peter, kremer}@mpip-mainz.mpg.de

In this lecture we give an overview of simulation approaches to soft matter systems. We explain important properties of soft matter, in particular of polymeric materials. We introduce different methods and levels of resolution in soft matter simulations, focusing on particle-based classical methods. Finally, we show the principles of multiscale simulation methods, the concept of scale bridging and methods used to systematically devise coarse grained simulation models.

1 Introduction – Soft Matter Systems

Soft matter systems – as opposed to hard matter such as minerals – are characterized by a comparatively low energy density and the relevant energy scale of the order of the thermal energy. As a consequence the properties of these materials are very much dominated by thermal fluctuations, i.e. entropic contributions. This importance of thermal fluctuations has a big impact on the molecular simulation methods that are appropriate for soft matter systems, as will be explained in detail below. Typical examples for soft matter are classical synthetic polymers, biological systems such as biopolymers and biological membranes, complex fluids, colloidal suspensions etc.

In the present lecture we focus mostly on theory and simulation of macromolecular systems, i.e. molecules that may contain many thousands of atoms. In this context classical synthetic polymers are a prototypical class of systems, as their molecular structure in most cases is simpler than those of typical biopolymers. In the simplest case a polymer is a long chain molecule of identical repeat units (beads or monomers), and the bulk (melt or solution) system constituted by these molecules is in the easiest case amorphous, homogeneous, and isotropic. Obviously this is a simplification and modern polymer chemistry has produced a variety of complex structures and materials. Nevertheless many characteristic problems and methods to solve them by computer simulations can be illustrated very well already for these simple (generic) model systems.

For this reason we structure this lecture as follows: first, we will briefly introduce a few general concepts of the statistics of dense polymeric systems to lay a foundation to later understand better the results and properties obtained in computer simulation of soft matter systems. In the subsequent section we will introduce the different simulation methods and scales used, again with a strong emphasis on models used in polymer simulation. In the last section we will show, how – in a multiscale simulation approach – the different simulation scales can be combined into a powerful tool that can address a variety of complex soft matter problems.

1.1 Polymers – chain statistics, scaling laws etc.

Dense polymer systems such as melts, glasses and crosslinked melts or solutions (networks such as rubber and gels) are very complex materials. Besides the local chemical interactions and density correlations, which are common to all liquids and disordered solids the global chain conformations and the chain connectivity play a decisive role for many physical properties. Local interactions determine the liquid structure on the scale of a few $\mathring{A}$ or at most a few nm. To study such local properties the chemical details of the chains have to be taken into account in the simulations and atomistically detailed melts are considered. However, if we look at the dynamics of a polymer chain in such a melt, local interactions do determine the packing and the bead friction but they affect generic properties only in a rather indirect way[1]. It is the main focus of the present contribution to discuss generic aspects common to all polymers and then later on go back to the question to what extent chemistry specific aspects play a role or make a difference. The consequences of the latter are also termed as structure property relations (SPR) in more applied research[1,2].

To stick to simple situations we consider polymer melts or networks where the chains are all identical. They can be characterized by an overall particle density ρ and a number of monomers N per chain. The overall extension of the chains is well characterized by the properties of random walks[3–5]. With ℓ being the average bond length we then have (for $N >> 1$) for the mean square end to end distance

$$\langle R^2(N)\rangle = \ell_K \ell (N-1) \approx \ell_K \ell N \tag{1}$$

and $\langle R_G^2(N)\rangle = \frac{1}{6}\langle R^2(N)\rangle$ for the radius of gyration, which is the mean squared distance of the beads from the center of mass of the chain, respectively [a]. ℓ_K is the Kuhn length and a measure for the stiffness of the chain. This gives an average volume each chain covers of

$$V \propto \langle R^2(N)\rangle^{3/2} \sim N^{3/2} \tag{2}$$

leading asymptotically to a vanishing self density of the chains in a melt. In order to pack beads at a monomer density ρ, which is a typical density of a molecular liquid, $0(N^{1/2})$ other chains share the very same volume of the chain and their conformations strongly interpenetrate each other. These other chains effectively screen the long range excluded volume interaction, since the individual monomer cannot distinguish, whether a non-bonded neighbor monomer belongs to the same chain or not. This leads to the above mentioned random walk structure, unlike dilute solutions, where the chains are more extended and display the so called self avoiding walk behavior with different scaling exponents[3]. This general property is firmly established by experiment and many simulations[6].

On length scales much larger than R_G^2 polymers diffuse as a whole and the motion is well described by standard diffusion. However over distances up to the order of the chain size, the motion of a polymer chain is more complex, even though hydrodynamic interactions are screened and do not play a role. On smaller length scales, the random diffusive motion of a monomer is constrained by the chain connectivity and the interaction with other monomers. To a very good first approximation, the other chains can be

[a] In dilute solution the situation is somewhat different. In the case of a good solvent the chains are more expanded and $\langle R^2\rangle \propto N^{2\nu}$ with ν close to 3/5

viewed as providing a viscous background and a heat bath. This certainly is a drastic oversimplification, which ignores all correlations due to the structure of the surrounding. The advantage of this simplification is that the Langevin dynamics of a single chain of point masses connected by harmonic springs can be solved exactly[7]. This was first done in a seminal paper by Rouse[8] and about the same time in a similar fashion by Bueche[9]. In this model, which is commonly referred to as the Rouse model, the diffusion constant of the chain $D \sim N^{-1}$, the longest relaxation time $\tau_d \sim N^2$ and the viscosity $\eta \sim N$. This describes the dynamics of a melt of relatively short chains, meaning molecular weights of e.g. $M \leq 20\,000$ for polystyrene [PS, M_{mon}=104] or $M \leq 2000$ for polyethylene [PE, M_{mon}=14], both qualitatively and quantitatively almost perfectly, though the reason is still not well understood. Only recently some deviations have been observed[10]. The effects are rather subtle and would require a detailed discussion beyond the scope of this lecture. For longer chains, the motion of the chains are observed to be significantly slower. Experiments show a dramatic decrease in the diffusion constant[11], $D \sim N^{-2.4}$, and an increase of the viscosity[7] towards $\eta \sim N^{3.4}$. The time-dependent elastic modulus $G(t)$ exhibits a solid or rubber-like plateau at intermediate times before decaying completely. Since the properties for all systems start to change at a chemistry- and temperature-dependent chain length N_e or molecular weight M_e in the same way, one is led to the idea that this can only originate from properties common to all chains, namely the chain connectivity and the fact that the chains cannot pass through each other.

Such dynamics as well as the previously mentioned static properties are the same for all polymers. This gives rise to the assumption that indeed the most simple polymer models, which are ideal test systems for simulations, should be sufficient to investigate these properties. Indeed, as was shown by de Gennes in a famous work on the relation between critical phenomena and macromolecules, one can view the inverse chain length $1/N$ as the normalized temperature distance $\frac{T-T_c}{T_c}$ from the critical point in a special ferromagnetic model ($n \to 0$ model). This means that for large N, i.e. close to T_c, all scaling properties and ratios of the related prefactors are universal and independent of chemical or model details. This finding is the theoretical foundation of the very successfully studied simple, generic models which we describe in the next section. Before that however, we first look more closely at local nonuniversal aspects.

2 Simulation of Soft Matter – A Wide Range of Resolution Levels

Properties and questions concerning soft matter systems cover a large range of length and time scales. Both local chemically specific interactions (e.g. specific attraction of certain units to surfaces, hydrogen bonding in aqueous environments and many, many more) as well as mesoscale effects such as hydrodynamic interactions or the formation of mesoscopic superstructures determine the behavior and the material properties of soft matter systems. For this reason, an equally wide range of simulation methods at different levels of resolution and consequently including a different amount of degrees of freedom is employed to study them. In this section we will give a very brief overview of the different individual simulation scales, show for which types of systems, questions, properties, length scales, time scales and types and number of degrees of freedom they are typically used. The basics of the methods we refer to in this chapter, namely molecular dynamics (MD) and monte carlo (MC) simulations, have been covered already in the lectures by

Sutmann and Kratzer.

2.1 QM methods

We will only very briefly linger with quantummechanical (QM) methods. Quantummechanical simulation approaches include electronic degrees of freedom, providing different levels of approximations to solve the Schrödinger equation. They provide the most detailed picture of the system, while being at the same time very limited in terms of system sizes and time scales available. These methods by themselves consist of a large variety of methodologies with very different levels of approximation and consequently accuracies concerning the electronic structures and energies provided. Overviews of several different QM approaches are given in other lectures at this school (Hättig, Zeller, Thiel, Reuter, etc.)

QM methods provide electronic energies of the systems studied, they are however limited in the sampling of thermal/statistical fluctuations, conformations etc. For this reason these methods are ideally suited for the simulation of hard matter systems such as minerals or metals. In the case of soft matter, where energy differences typically are in the order of thermal energies and consequently fluctuations, statistical sampling and entropy arguments play a decisive role, QM methods alone are often not sufficient to study structures and properties of the respective systems. However, QM methods are extremely valuable to provide interaction energies etc. as parameters for interaction functions in classical simulation methods (see below). In addition, QM methods are frequently used in mixed QM/MM (quantum/classical) approaches.

2.2 Classical atomistic simulations

Classical atomistic simulation models do not include electronic degrees of freedom. In the corresponding approaches atoms are treated as classical particles that interact via a set of interaction potentials called a forcefield. Typically these forcefields consist of intramolecular, covalent (also often denoted as bonded) interactions, i.e. corresponding to bonds, angles, torsions etc and nonbonded interactions (including also electrostatics). Forcefield parameters are usually determined either by making use of quantummechanical reference calculations (frequently used for example to determine parameters for bonded potentials and atomistic partial charges), or by parameterization such that experimental data (densities, diffusion constants, dielectric properties, solvation thermodynamics etc.) are reproduced[12–16].

In the case of soft matter simulations, classical atomistic atomistic models are typically used in cases, where specific interactions play an important role, for example in biomolecular systems[17], in cases where explicit solvent degrees of freedom are important (this applies of course also for biomolecules), or for example in the case of liquid mixtures (in particular if one wants to compute thermochemical data such as free energies for such systems[18]), but also if one wants to study local monomer packing or local dynamics.

2.3 Coarse grained particle-based models

In the introduction the interplay of different length and time scales has been mentioned. In order to make a quantitative comparison to specific experiments in most cases one has

to be both chemistry specific and asymptotic (scaling) aspects are equally important and have to be considered accordingly. The combination of both aspects will be discussed in the section on *Multiscale Simulation*. This led to a powerful methodology, which now can be used for rather different polymer species and the systematic extension to complicated macromolecules such as multiblock copolymers, gels or proteins is an important current research topic. On the other hand both force filed simulations or rather small systems and rather short times or the simulation of most simple generic models can by itself contribute in many ways to our understanding of material properties, as has been illustrated for the case of force fields in the previous section of this chapter.

Simple generic models are perfectly suited to study scaling properties of macromolecular systems as they reduce the computational complexity to the absolute minimum, namely connectivity and excluded volume plus some specific interactions, if needed. This allows for extremely efficient algorithms allowing for much longer effective time and lengths scales and much smaller error bars than in more detailed simulations. This leads to a first important message of this chapter, namely that one generally should use the most simple model, which is able to capture the essential physics and chemistry under study. Though this statement is not new at all, this still holds even for most modern computers.

There are many different possibilities to study idealized particle-based polymer models on a computer. They range from simple lattice models, which can be treated by a variety of different Monte Carlo methods to continuum models. Continuum models have both been studied by Monte Carlo and Molecular Dynamics simulations. Especially for questions not requiring a high particle density the stochastic Monte Carlo method can be computationally advantageous and has been used extensively to study both static and dynamic properties. In the present contribution we confine ourselves to Molecular Dynamics simulations. For a comparison to MC we refer refer the reader to the literature[19,20].

2.3.1 Bead-spring models

Models where the individual polymer chain is modelled by mass points which repel each other and which are connected along the chain by a spring are the most elementary MD models. The repulsion produces the excluded volume constraint and the spring takes care of the connectivity.

A quite commonly used model is that of Kremer and Grest[21–23]. We consider beads with a unit mass. All beads interact via a purely repulsive Lennard-Jones (WCA Weeks Chandler Anderson) potential, to model the excluded volume interaction.

$$U_{LJ}^{r} = \begin{cases} 4\epsilon \left\{ (\sigma/r)^{12} - (\sigma/r)^{6} + \frac{1}{4} \right\} & r \leq r_c \\ 0 & r \geq r_c \end{cases} \tag{3}$$

with a cutoff $r_c = 2^{1/6}\sigma$. The beads are connected by a finite extensible non-linear elastic potential (FENE)

$$U_{FENE}^{(r)} = \begin{cases} -0.5 R_o^2 k \ell_n (1 - (r/R_o)^2) & r \leq R_o \\ \infty & r > R_o \end{cases} \tag{4}$$

in addition to the Lennard Jones Potential. In melt simulations the parameters are usually taken as $k = 30\epsilon/\sigma^2$, $R_o = 1.5\sigma$. This choice of parameters ensures, that the chains

never cut through each other in the course of a simulation. The temperature $T = \epsilon/k_B$ and the basic unit of time is $\tau = \sigma(m/\epsilon)^{1/2}$. Volume and temperature are usually kept constant. The temperature is kept constant by coupling the motion to a Langevin thermostat[24]. An alternative, which does not screen hydrodynamics would be a DPD (dissipative particle dynamics) thermostat. The average bond length with these parameters is $\langle \ell^2 \rangle^{1/2} = 0.97\sigma$. In addition a bending stiffness can be introduced via an effective three body interaction. With $cos\Theta_i = (\hat{\mathbf{r}}_{i,i-1} \cdot \hat{\mathbf{r}}_{i,i+1})$, with $\hat{\mathbf{r}}_{i,i-1} = (\mathbf{r}_i - \mathbf{r}_{i-1})/ \mid \mathbf{r}_i - \mathbf{r}_{i-1} \mid$ and accordingly $\hat{\mathbf{r}}_{i,i+1}$ the bending potential reads $U_{bend}(\Theta) = k_\Theta(1 - cos\Theta)$. By variation of k_Θ the model can be tuned from very flexible to very stiff.

While models along this line have also been employed for solutions, the main area of application is the study of dense polymer melts, mixtures and networks. For this model in a melt of density $\rho = 0.85\sigma^{-3}$ with $k_\Theta = 0$ one finds $c_\infty l^2 = ll_k = 1.7\sigma^2$. With the parameters given above a chain crossing is essentially impossible, i.e. forbidden by a Boltzmann factor of about $exp(-70)$, which means that it never happens during a simulation. Thus crucial problems as illustrated in Fig 1 can be easily studied.

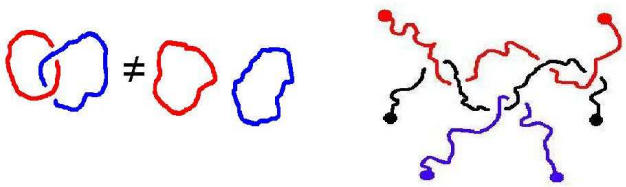


Figure 1. (a) Illustration of two topologically different states of rings, which are very difficult to separate analytically when calculating partition functions of an ensemble. (b) Illustration of "many chain" effects for topological constraints in a melt or network. Taking out the blue chain also releases all constraints between the red and the black chain.

Since in these models there are no torsional barriers, the monomer packing relaxes quickly and the simulations is very efficient. The packing locally depends very sensitively on the ratio of bond length to effective excluded volume of the beads[25]. However the local equilibration of the sphere packing is not sufficient for the overall equilibrium. The local packing only characterizes an equilibration on the smallest length scale considered and very small systematic deviations on that scale can lead to significant deviations from equilibrium at the large length scales on the order of the chain extension. We thus have to ensure proper equilibration on all lengths scales. One way would be to run a conventional MD simulation until all length scales are properly relaxed and in equilibrium.This however however quite often requires a prohibitively large amount of computer time. A well working strategy for many model systems is as follows[23]:

- Simulate a melt of many short, but long enough chains ($N\ell \gg \ell_K$) into equilibrium by a conventional method

- Use this melt to construct the master curve or target function for the melts of longer chains of all internal distances.

- Create non reversal random walks[b] of the correct bond length, which match the target function closely. They have to have the anticipated large distance conformations. Introduce, if needed beyond the intrinsic stiffness of the bonds, stiffness via a suitable second neighbor excluded volume potential along the chain. (This might be a bit larger than the one of the full melt!)

- Place these chains as rigid objects in the system and move them around by a Monte Carlo procedure (shifting, rotating, inverting..., but **not** manipulating the conformation itself) to minimize density fluctuations

- Use this state as starting state for melt simulations

- Introduce slowly but not too slowly the excluded volume potential by keeping the short range intra chain interactions, taking care that in the beginning the chains can easily cross through each other

- Run until the excluded volume potential is completely introduced. Control internal distances permanently to check for possible overshoots, deviations from the master curve.

- Eventually support long range relaxation by so called end bridging[26,27] or double pivot moves[23]

Independent on the details of the procedures, it is important to continuously monitor the actual structure to the master or target curves. If one is stuck with this approach and deviations from the master curve occur and stay, one has to start all over again. Ref.[23] also demonstrates some typical deviations from the master curve as they can occur and which should be avoided during the set up.

Following such a procedure for melts or at least monitoring the internal distances i.e. for networks assures that the simulations start out from well defined equilibrium states.

These models have been used quite frequently to study the dynamics of short and long chain polymer melts as well as relaxation properties of elastomers, cross linked polymer melts. For the latter we just illustrate the importance of the noncrossability for an ideal model network. The network is built like a diamond crystal, where the bonds are replaced by polymer chains. This has the advantage that many properties within the classical rubber elasticity models can be calculated exactly. In order to obtain network strands of zero or at least low tension, they have to obey random walk statistics, just like corresponding chains in a melt. This means that such a diamond lattice network is not space filling and thus we need randomly interpenetrating networks where the strands are random walk like and simultaneously melt density is reached. As a consequence the only source of frozen disorder comes from random links between network loops of different sublattices. A direct way to visualize this is to stretch such a network and ”measure” the tension of the strands. While this can be done quantitatively to demonstrate the relevance of entanglements for networks, we here stick to a graphical illustration. Fig. (2) gives the result of such a visual inspection of the stress distribution in a network. Those network strands, which are

[b] Non reversal random walks are RWs, which do not immediately fold back in themselves. By this the length scale correlations along the backbone of the chains are properly reproduced even though there is no excluded volume

linked topologically to other strands, lead to shorter paths connecting through the whole network. These short paths carry most of the stress under strong elongation and are the first to break. In experiment of course a combination of short chemical and topological connections through the network exist. Their breaking is a reason for the fact that in a typical elastomer the elastic modulus after the first strong extension is a bit smaller than right after the crosslinking reaction.

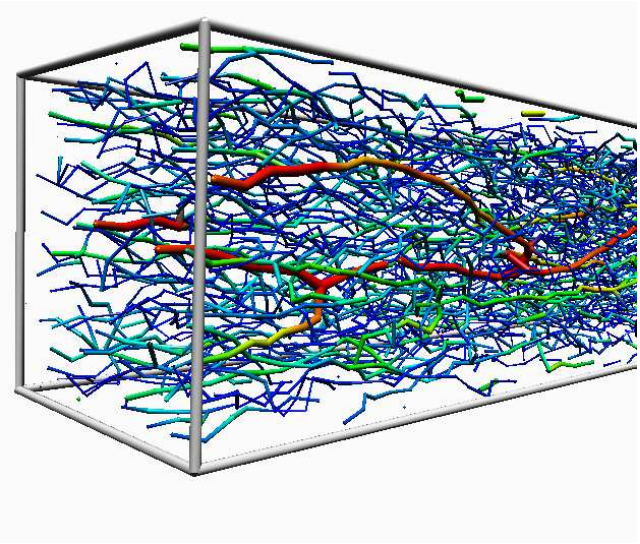


Figure 2. Elongated diamond lattice networks, where the only source of disorder are random links between network loops. The strands are shown due to their stretching (similar to the stress they carry) from small stretching (thin) to strong stretching (thick). From Ref[28]

Starting out from a polymer network, one can ask the question, what would happen, if the strands N_S between crosslinks are made extremely large. Experiments show that the elastic modulus of the network decreases as $G^0(N_S) \propto 1/N_S$, but only up to a characteristic length, which varies with chemistry. For longer chains the modulus stays constant. The second important observation is that the time dependent modulus $G(t)$ first decays like in a liquid and then levels of to $G^0(N_S)$. This observation actually also holds for melts without any crosslinks over many orders of magnitude in time as well, if the chains are much longer than a characteristic chain length N_e or M_e, the entanglement molecular weight. The longest relaxation time in such a melt scales as $\tau_d \propto N^{3.4}$, the disengagement time[c]. This time can be macroscopically large that a melt is almost indistinguishable from a network. Fig. (3) illustrates the idea which explains this phenomenon in a way, that for very long chains inner parts for intermediate times do not "know" whether the ends are crosslinked or not. As a consequence the chains have to diffuse along their own contour and the lateral movement is constrained to an effective tube. This is the essence of the so called reptation model of deGennes and Edwards. The backbone of the tube is called

[c]The theoretical prediction for the asymptotic behavior of very long chains actually is $\tau_d \propto N^3/{N_e}^7$

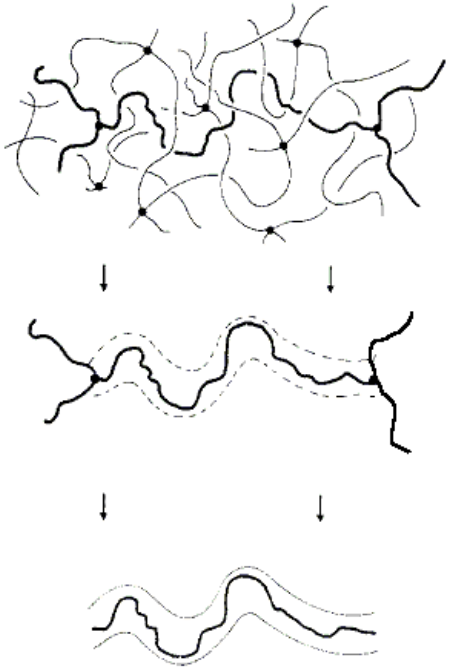


Figure 3. Sketch of the historical development of the tube constraint and reptation concept. Starting from a network Edwards in 1967 defined the confinement to the tube, while deGennes in 1971 realized that for long chains the ends only play a small role for intermediate times.

primitive path and the entanglement length N_e corresponds to an unperturbed melt chain, which just fits into the tube. This model and its consequences have been studied extensively by theory, simulation and experiment and it is by now well established[29,30]. Despite this effort it was very difficult to predict N_e from the chemical structure or static properties of the polymers. Here computer simulations offer an interesting approach by determining the contour length of the primitive path L_{PP}. This can be achieved in a simulation by contracting the chains as much as possible, while keeping the end to end distance fixed. By this all existing constraints are conserved, while simultaneously many chain effects are automatically taken into account. This actually is a modern version of the original idea of Edwards from 1967[31,32], who defined the primitive path in a network as the shortest path between the endpoints of a chain, which does not violate any topological constraint. With the ansatz

$$L_{PP} = a_{PP} N / N_e \tag{5}$$

$$a_{PP} = (l_k N_e)^{1/2} \tag{6}$$

N_e can directly be read off the simulation result. a_{PP} is the step length of the primitive path. This leads to a direct prediction of one of the central rheological quantities for polymer melts and networks just from the conformational properties. The outcome of such simulations can now be compared to experiment via a recent scaling theory. If the lo-

cal packing of the chains leads to the primitive path, it should be possible to estimate the plateau modulus G^0 of such a melt as a function of the length scales, which determine the overall conformations. These two lengths scales are the Kuhn length of the chains, l_k, as introduced before, and the packing length $p = N/(\rho\langle R^2(N)\rangle)$. While l_k is a measure for the chain stiffness, p gives the typical lateral distance between the strands[d]. ρ is the bead density in the simulation. With this information and the expression for the modulus based on the reptation theory

$$G^0(N_e) = \frac{4}{5}\frac{\rho k_B T}{N_e} \tag{7}$$

we can plot the dimensionless modulus $G^0(N_e)k_bT/l_k^3$ versus the ratio l_k/p. Here, k_bT is the thermal energy which is the natural energy scale for soft matter. The result is shown in Fig. (4).[33–35]

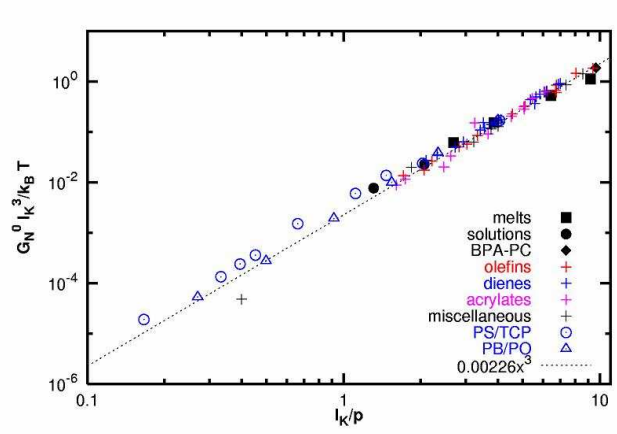


Figure 4. Dimensionless plateau moduli $G_N^0 l_K^3/k_BT$ as a function of the dimensionless ratio l_K/p of Kuhn length l_K and packing length p[33,34]. The figure contains (i) experimentally measured plateau moduli for polymer melts[36] ($*$ Polyolefins, $\times$ Polydienes, + Polyacrylates and $\triangleright$ miscellaneous), (ii) plateau moduli inferred from the normal tensions measured in computer simulation of bead-spring melts[37] ($\square$) and a semi-atomistic polycarbonate melt[38] ($\diamond$) under an elongational strain, and (iii) predictions of the tube model Eq. 7 based on the results of our PPA for bead-spring melts ($\blacksquare$), and the semi-atomistic polycarbonate melt ($\blacklozenge$). The line indicates the best fit to the experimental melt data for polymer melts by Fetters et.al [39]. Errors for all the simulation data are smaller than the symbol size.

The agreement is excellent showing the overall consistency of the previous discussion. Meanwhile Everaers and coworkers have extended this kind of analysis to semi flexible (bio)polymers such as dense solutions of DNA or actin filaments, where the random walk assumption of the for the chain on the scale of the tube diameter does not hold anymore. There extension now gives a consistent picture in the scaled modulus for 12 decades![35]

[d]Actually these two lengths are not independent of each other. For present discussion however we stick to the assumption that they can be treated as independent. Also there are some ambiguities in the definition of l_k in entangled solutions, which we do not discuss here

Current research in this field, besides the topics discussed below, mostly deals either with the problem of tube deformation under shear, gels of semi flexible polymers, branched systems and situations, where one tries to temporary prevent chain entanglements in order to improve the processibility of such materials, the latter being very important for many every day plastics but also high tech applications.

2.4 Beyond (purely) particle-based

In the case of (semi-)dilute polymer solutions hydrodynamic effects play a major role, which are completely screened in dense polymer systems. Hydrodynamic interactions act on a mesoscopic length and time scale which is hardly accessible to the purely particle based approaches discussed above. The corresponding theoretical concepts and numerical methods (for example the Lattice Boltzmann method, where the solvent fluid is not represented by explicit particles by via a grid) are discussed in a separate lecture by B. Dünweg.

Also for the computation of many material properties such as mechanical properties of composite materials, particle-based methods provide a too microscopic picture. For these properties, other numerical methods (continuum representations which for example make use of finite element solvers) have been developed (see for example the lecture of ...).

3 Multiscale Simulation / Systematic Development of Coarse Grained Models

”Multiscale simulation” refers to methods where different simulation hierarchies are combined and linked to obtain an approach that simultaneously addresses phenomena or properties of a given system at several levels of resolution and consequently on several time and length scales. Multiscale simulation approaches can operate in different ways of combining the individual levels of resolution: the easiest way to combine different simulation models on different scales is to treat them separately and sequentially by simply passing information (structures, parameters, energies etc.) from one level of resolution to the next. In the case of hybrid simulations, different levels of resolution are present simultaneously. This is more complex than the sequential approach since interactions between particles/entities at different levels of resolutions have to be devised. An even more sophisticated multiscale approach allows to adaptively switch between resolution levels for individual molecules on the fly – for example depending on their spatial coordinates. This is more complex since the exchange needs to be carefully conducted to adhere/conserve statistical mechanical principles (conservation laws, prevent fluxes, etc.). For details on adaptive resolution multiscale methods see the lecture by Delle Site and Junghans.

Irrespective of the method to combine the individual scales it is an important property of a ”true” multiscale simulation approach that the individual models on different levels of resolution are systematically linked. This scale bridging requires systematic development of the individual models such that they are thermodynamically and structurally consistent.[40–46],

Many different approaches have been followed to obtain systematically linked simulation models on different levels of resolution, both from the QM to the classical level and from the classical all-atom level to a coarse grained description. In particular for the latter

case, one can distinguish between several distinct approaches: in the energy-based coarse graining approach the interaction potentials between the coarse grained particles are derived such that energies or free energies obtained at the atomistic level are reproduced[40,46]. In the force matching method, the forces in the CG system are determined such that they are mapped to the sum of the forces on the corresponding atomistic system[43]. The structure-based CG methods provide CG interactions that reproduce a pre-defined target structure – often described by a set of radial distributions functions obtained from all-atom molecular simulations[47–49]. Note that there is currently much research being carried out to investigate, whether – and if yes how – it is possible to derive coarse grained potentials that are both thermodynamically as well as structurally consistent with the underlying higher-resolution description.

In the following we will focus on the structure-based coarse graining methodology which was originally developed in the field of polymer simulation. We will illustrate the different steps that need to be taken to develop a coarse grained model (most of which are in spirit of course not limited to structure-based coarse graining), we will show examples both from the original scope of amorphous polymer melts to more complex systems, where chemically specific/thermodynamic aspects start to play an increasing role.

The subsequent section is organized along the sequence of steps in the coarse graining process: first, a mapping scheme is formulated that relates the coordinates in the atomistic description with the coarse grained ones. Second, one has to decide on a strategy concerning bonded and nonbonded interactions. This distinction is based on the assumption that the total potential energy U^{CG} can be separated into bonded/covalent and nonbonded contributions

$$U^{CG} = \sum U_B^{CG} + \sum U_{NB}^{CG} \,. \tag{8}$$

In the methodology followed here, non-bonded (intermolecular) and bonded (covalent, intramolecular) interactions are separated as clearly as possible and derived sequentially. In particular for nonbonded interactions, several approaches have been developed to derive interaction potentials between coarse grained particles. We focus here on the structure-based coarse graining methodology which will be discussed in detail below. After one has obtained mesoscale structures and long-time trajectories of the CG system by MD simulations, the last step consists of reintroducing atomistic details ("back-mapping" or inverse mapping) into the CG simulation trajectory. This also belongs to the coarse graining procedure in the sense that it provides a crucial link between the atomistic and the coarse grained level of resolution.

In the course of the following sections, we will illustrate the above steps along several examples: (i) BPAPC (only briefly)[25], and (ii) polystyrene being typical examples for amorphous polymeric systems[50,51]. With (iii) the low molecular weight liquid crystalline compound 8AB8 we show how the recipes from the polymer world can be extended to systems, where chemically specific nonbonded interactions play an increasingly important role[49]. As a last step we will give with (iv) a dipeptide (diphenylalanine) in aqueous environment an outlook to show how the methodology can be extended to systems where the complexity compared to homogeneous isotropic polymer melts is increased[52,53]. Figure 5 shows the chemical structure of the named compounds.

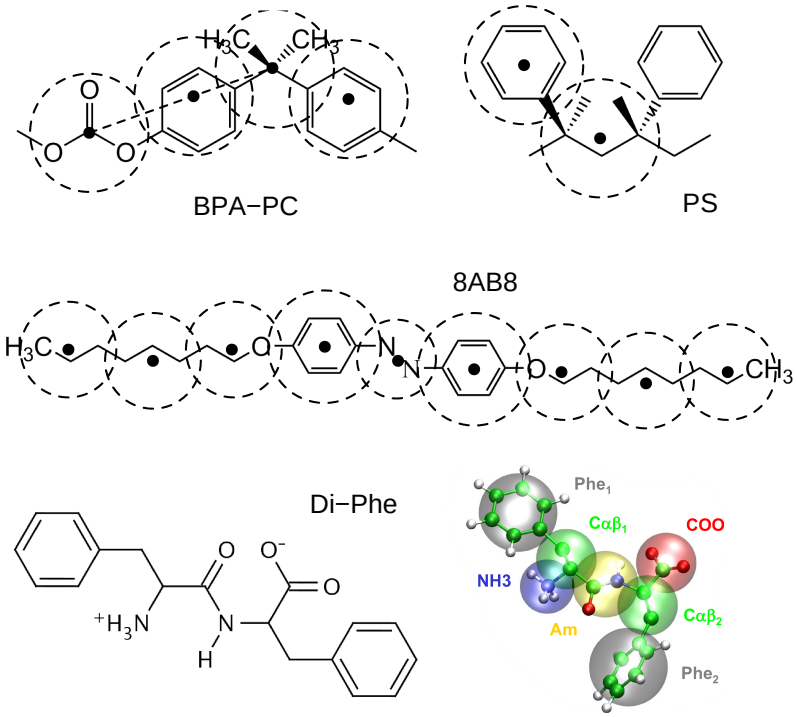


Figure 5. Chemical structure and mapping schemes of the discussed CG examples: BPA-PC, Polystyrene (PS), 8AB8, diphenylalanine (Di-Phe)

3.1 Mapping scheme

The mapping scheme relates the atomistic coordinates of a structure with the bead positions in the CG model. Here, we only consider CG centers with spherically isotropic potentials. Even though there is no unique way to map a given set of atoms onto a coarser description[25], there exist empirical criteria for a good choice of mapping points which depend on the specific properties under investigation. A very important criterion for a mapping scheme is its ability to decouple internal degrees of freedom so that the intramolecular (bonded) potentials can be cleanly factorized into bond, angle and torsion terms. For example for the liquid crystalline compound 8AB8, which can undergo a transition from a rod-like extended (*trans*) structure to a V-shaped bent (*cis*) structure, one needs a clear distinction between the two states since they are crucial for the phase behavior of the compound. The most convenient way to capture the geometry change of the AB unit upon photoisomerization is to position a "linker bead" in the center of the azo group. Consequently, the centers of the two phenyl groups appear to be logical next CG centers. With these CG phenyl beads we can later on also compare the applicability of universally (re-)usable CG phenyl parameters based on liquid benzene with parameters derived specifically for azobenzene.

Another example for the importance of the mapping scheme chosen is BPAPC. Here

the at first counterintuitive effect occurs that a coarser mapping scheme which reduces the number of beads in the system can nevertheless effectively lower the computational efficiency compared to a scheme with more CG particles. This is because in the latter case the polymer appears to be more *tube like* which leads to a reduced friction in the system and effectively speeds up the dynamics in the simulation[54].

3.2 Bonded interaction potentials

Bonded interactions derived such that the conformational statistics of a single molecule is represented correctly in the CG model. The general strategy is to use Boltzmann inversion to convert distributions of conformational properties such as interparticle distances or angles into potentials.

Intramolecular bonded interactions of the CG model are determined by sampling the distributions of (CG) conformational degrees of freedom of an isolated molecule in atomistic resolution (determined here by MD simulations using a stochastic thermostat to ensure proper equilibration). These conformational distributions P^{CG} are in general characterized by specific CG bond lengths r between any pair of CG beads, angles θ between any triple of beads and torsions ϕ between any quadruple of beads respectively, i.e. $P^{CG}(r, \theta, \phi, T)$. If one assumes that the different CG internal degrees of freedom are uncorrelated, $P^{CG}(r, \theta, \phi, T)$ factorizes into independent probability distributions of bond, angle and torsional degrees of freedom

$$P^{CG}(r, \theta, \phi, T) = P^{CG}(r, T)\ P^{CG}(\theta, T)\ P^{CG}(\phi, T)\ . \tag{9}$$

This assumption has to be carefully checked (it is not uncommon that coarse grained DOFs are correlated, for example that certain combinations of CG bonds, angles and torsions are "forbidden" in the distributions obtained from the "real" atomistic chain), and is an important test of the suitability of a mapping scheme. A mapping scheme containing complex multi-parameter potentials is computationally rather inefficient. The individual probability distributions $P^{CG}(r, T)$, $P^{CG}(\theta, T)$, and $P^{CG}(\phi, T)$ are then Boltzmann inverted to obtain the corresponding potentials:

$$U^{CG}(r, T) = -k_B T\ \ln\left(P^{CG}(r, T)/r^2\right) + C_r \tag{10}$$

$$U^{CG}(\theta, T) = -k_B T\ \ln\left(P^{CG}(\theta, T)/\sin(\theta)\right) + C_\theta \tag{11}$$

$$U^{CG}(\phi, T) = -k_B T\ \ln\ P^{CG}(\phi, T) + C_\phi\ . \tag{12}$$

Note that these potentials are in fact potentials of mean force, ergo free energies and consequently temperature dependent (not only due to the prefactor $k_B T$). This means they can strictly be only applied at the temperature (state point) they were derived at, requiring a reparametrization at each temperature. In practice, one needs to test the width of this applicability range for each CG model individually. Experience shows that typical temperature ranges are of the order of $\pm 10\%$ (if no phase transition is within that range). Technically, the Boltzmann inversions (equations 10-12) and the subsequent determination of the derivatives can be carried out numerically, resulting in tabulated potentials and forces. Another option is to determine analytical potentials that reproduce the probability distributions $P^{CG}(r, T)$, $P^{CG}(\theta, T)$, and $P^{CG}(\phi, T)$, for example by fitting the (multi-peaked) bond and angle distributions by a series of Gaussian functions which can then be

inverted analytically resulting in smooth potentials and forces[55]. This latter method was used for 8AB8.

The approach of strictly separating into bonded and nonbonded CG potentials as outlined here is in contrast to other approaches, where the CG internal degrees of freedom are determined based on the distributions obtained from an atomistic simulation of the melt/liquid phase[48]. In the latter case one obtains potentials for bonded and nonbonded interactions simultaneously based on the same melt, consequently they are interdependent, i.e. there is no clear separation between covalent and nonbonded interaction potentials. For amorphous polymeric systems, we achieve this separation by deriving bond, angle and torsional distributions from the conformations of the single chain in vacuo. When generating these distributions, the inclusion of nonbonded interactions has to be taken with some care to avoid "double counting" of long range intra-chain interactions. For example in the case of 8AB8, we simulated a single 8AB8 molecule with MD using a stochastic thermostat and we excluded all intra-chain nonbonded interactions beyond the distance of three CG bonds (i.e. $\equiv$ a torsion), since the interaction between these beads will be covered by non-bonded interactions in the CG simulations. Note that in analogy short range bead-bead interactions along a chain are covered by bond, angle and torsion potentials, which means for these bead pairs nonbonded interactions need to be excluded in the CG simulations.

The above clear separation by simulation of a single molecule in vacuo can only be successful, it the conformational sampling of the molecule in vacuo and in the bulk (or solution) phase does not differ substantially. In biomolecular systems due to the peculiar nature of aqueous solutions (i.e. the presence of hydrogen bonds) this assumption gets problematic, as can be seen for the dipeptide diphenylalanine[52].

3.3 Nonbonded interaction potentials

In the structure-based nonbonded interaction potentials between coarse grained beads are derived based on the structure of isotropic liquids of small molecules (in the case of more complex molecules such as 8AB8, fragments of the target molecule are used).

In structure-based coarse graining approaches nonbonded interaction potentials are derived such that structural properties of the liquid or melt are reproduced. In this case, radial distribution functions of the atomistically simulated liquids are used as targets for the parametrization process. For the CG potentials and their optimization there are two principal options: the first option is to adjust the parameters of analytical potentials such as Lennard-Jones to closely reproduce the structure of the atomistic melt/liquid. The second option is to use numerically derived tabulated potentials which are designed such that the CG melt exactly reproduces the atomistic melt structure.

In the first case the standard Lennard-Jones 12-6 potential has turned out to be too strongly repulsive, i.e. too "hard", for CG particles which are rather large and soft. Softer Lennard-Jones-type of potentials (e.g. 9-6 or 7-6[51]), or Morse potentials have proven to be more useful. In the present study, we used Morse potentials

$$U^{CG}_{NB,morse}(r) = \epsilon \, (1 - e^{-\alpha \, (r-\sigma_M)})^2 \;\; \text{for } r < \sigma_M \quad \text{and} \quad 0 \;\; \text{for } r \geq \sigma_M \; . \qquad (13)$$

In the present form they are purely repulsive potentials in the spirit of the Weeks-Chandler-Andersen (WCA) model[56], shifted upwards and truncated in the minimum at $r = \sigma_M$. To adjust the parameters a simplex algorithm can be used[57,58] to search the optimal set of parameters to reproduce a given melt structure.

In the second scheme, the iterative Boltzmann inversion method[47,59] can be used to generate numerically a tabulated potential that exactly reproduces a given melt structure, i.e. a given radial distribution function $g(r)$. This method relies on an initial guess for a nonbonded potential $U^{CG}_{NB,0}$. Usually the Boltzmann inverse of the target $g(r)$, i.e. the potential of mean force,

$$U^{CG}_{NB,0} = -k_B T \ln g(r) \ , \tag{14}$$

is used, with which one then performs a coarse grained simulation of the liquid. The resulting structure of this first step will not match the target structure as the potential of mean force is – due to multibody interactions – only in the limit of very high dilution a good estimate for the potential energy. However, using the following iteration scheme

$$U^{CG}_{NB,i+1} = U^{CG}_{NB,i} + k_B T \ \ln(\frac{g_i(r)}{g(r)}) \ , \tag{15}$$

the original guess can be self-consistently refined until the desired structure is obtained.

For complex molecules with a large number of different CG beads or more importantly in the case of liquid crystalline molecules with anisotropic structures the procedure to determine nonbonded interaction functions is more complicated. In these cases it is advantageous to split the target molecule into fragments so that the nonbonded interactions between different bead types can be determined based on the structure of *isotropic* liquids of these fragment molecules. A potential error that could be introduced by this fragment-based approach is that different conformations or relative orientations between molecules might contribute differently to the structure of the fragment liquids than to the target liquid or melt. Consequently these conformations would be misrepresented in the CG potentials. One example are relative weights of parallel and perpendicular orientations between anisotropic molecules such as phenyl rings, that might differ in liquid benzene compared to molecules where the rings are embedded into a longer chain. This and other aspects of transferability of CG potentials[60,61], for example transferability between different compositions of liquid/liquid mixtures or the transferability of CG potentials between the *trans* and *cis* isomers of azobenzene need to be tested and will be more thoroughly discussed in the Results section. After careful testing and keeping in mind, that in principle all CG potentials are state dependent, the procedure to derive them from chain fragments and low molecular weight liquids opens up the possibility to reuse certain CG potentials for reoccurring building blocks (such as alkyl or phenyl groups).

3.4 Reintroduction of atomistic details ("back-mapping")

The "back-mapping" or inverse mapping problem, i.e. the task to reintroduce atomistic coordinates into coarse grained structures, has in general no unique solution since every CG structure corresponds to an ensemble of atomistic microstates. Thus, the task is to find a representative structure from this ensemble which displays the correct statistical weight of those degrees of freedom not resolved in the CG description, e.g. ring flips etc. Multiple strategies to reintroduce atomistic details into a CG structure have been presented[62,63,50]. The strategy was first to use reasonably rigid all-atom chain fragments – corresponding to a single or a small set of CG beads – which were taken from a correctly sampled distribution of all-atom chain structures. These fragments were fitted onto the CG structure,

the resulting all-atom structure was relaxed by energy minimization and a short equilibration[62,63,50]. In the case of more flexible low-molecular weight molecules (e.g. 8AB8 or diphenylalanine) a slightly different strategy was chosen to avoid the atomistic structure to deviate too strongly from the CG reference. Atomistic coordinates were inserted into the CG structure (either using presampled fragments or random coordinates) such that the constraint was applied that the atomistic coordinates have to satisfy the "mapping condition", i.e. the atomistic coordinates have to yield back the CG structure if one applied the mapping scheme. The resulting structure was then relaxed (energy minimized and equilibrated by molecular dynamics simulations), while at the same time always applying this "mapping condition". Practically, this was done by restraining the atomistic coordinates (with an additional external potential) to the CG mapping points. This procedure results in a perfectly equilibrated set of atomistic coordinates that almost (depending on the strength of the restraining potential) exactly reproduces the CG structure.

The combination of CG simulation with an efficient backmapping methodology is a powerful tool to efficiently simulate long time-scale and large length-scale soft matter processes where in the end one can obtain well-equilibrated atomistic structures. The resulting structures can be directly compared to experimental data[63] or they can be used in further computations, for example to determine free-energy data (e.g. the permeabilities of small molecules in large polymeric systems[64]).

Additionally, the combination of coarse grained simulations with a CG model based on an underlying atomistic description with a backmapping procedure can be further employed to validate the atomistic forcefield – on time and length scales not accessible to atomistic simulations alone due to sampling problems.

3.5 Time mapping

Within CG models length scales are usually well defined through the construction of the coarse graining itself. In most dynamic CG simulations reported in the literature little attention is paid however to the corresponding "coarse graining" of the time unit. From polymer simulations of both simple continuum as well as lattice models it is known that such simulations reproduce the essential generic features of polymer dynamics; that is, the crossover from the Rouse to the entangled reptation regime, qualitatively and to a certain extent quantitatively[29]. While such previous studies concern motion distances on scales well above a typical monomer extension and provide quantitative information on characteristic time ratios, this still leaves a number of open questions. These refer to the predictive quantitative modeling of diffusion, viscosity, rates, and correlation times, etc. of dynamic events as well as to the question of minimal time and length scales CG simulations apply to. Particle mass, size, and energy scale, which are all well defined within a CG model, of course trivially fix a time scale, and it is indeed this time scale that is most often reported in MD simulations of CG systems. However, it does not usually correspond to the true physical time scale of the underlying chemical model, because part of the friction experienced by a (sub)molecule (in the atomistic representation) is lost in the CG representation, causing the CG system to evolve faster. (Note that this is in principle also the case for atomistic simulations that make use of so-called united atoms where aliphatic hydrogen atoms are incorporated into the carbon atoms.) In other words, the fluctuating random forces of atomic DOFs, which are integrated out in the CG model, contribute to a "background friction" that

must be considered in order to obtain a realistic time scale in the CG dynamics simulation. The physical origin of the dynamic speedup in comparison with all-atom models and real-life experimental systems is that the barriers (e.g. for diffusional motion) are lower because CG interparticle potentials are softer and more smoothly varying with distance.

In order to determine the speedup in CG simulations due to these enhanced dynamics, CG dynamic quantities can in some cases be mapped directly onto the corresponding quantity obtained from detailed MD simulations or from experiments. For example, a diffusion coefficient in the coarse system D_{CG} can be mapped onto the diffusion coefficient in the atomistic system D_{AT}, effectively introducing a dimensionless scaling constant between the CG time unit and the actual time unit of the chemical system. This scaling factor can then also be used to estimate the actual speedup factor which for the present systems is about 10^4. Alternatively, the CG mean squared displacement curve can be superimposed with the atomistic curve at (for atomistic simulations) long times. This approach was used to study entangled polycarbonate (BPA-PC) melts of up to 20 entanglement lengths. The CG simulations provided truly quantitative information on the different measures of the entanglement molecular weight (from displacements, scattering functions, modulus and topological analysis) and the ratios of the different crossover times.

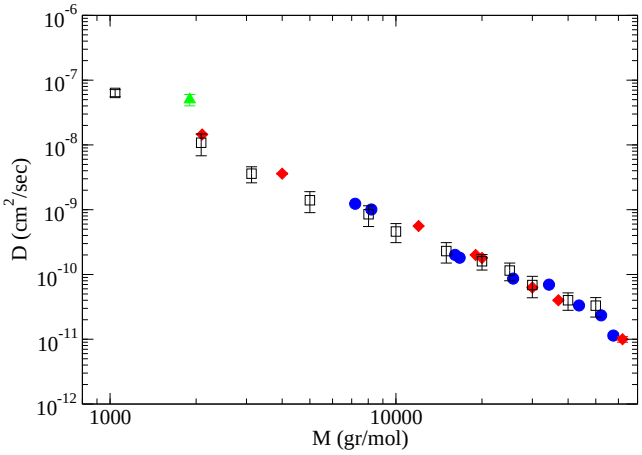


Figure 6. Diffusion constant of polystyrene chains in a melt of identical chains for molecular weights between $M_W = 1000$ and about $M_W = 50000$. The simulation data, obtained from a hierarchy of all atom, united atom and coarse grained simulation, do not contain any fitting parameter. The full symbols are from different experiments. For details see[65].

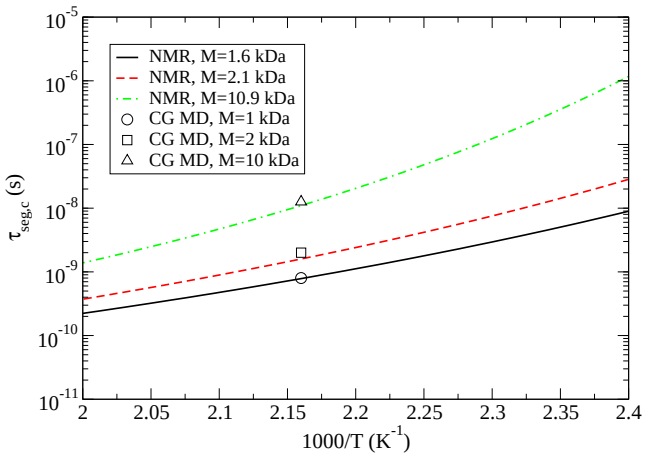


Figure 7. Segmental relaxation times of polystyrene as measured by NMR spectroscopy for different molecular weights as a function of temperature. Simulation data from CG runs for one temperature where the timescaling is the same as for the previous figure demonstrating the overall consistency of the approach [66].

Acknowledgments

We would like to thank Ralf Everaers, Gary Grest, Cameron Abrams, Luigi Delle Site, Vagelis Harmandaris, Nico van der Vegt, Berk Hess, and Alessandra Villa for many fruitful collaborations and stimulating discussions.

References

1. C. F. Abrams and K. Kremer, J. Chem. Phys., **116**, 3162, 2002.
2. K. Kremer, *Soft and Fragile Matter, Nonequilibrium Dynamics, Metastability and Flow*, NATO ASI Workshop, St. Andrews, 2000.
3. P. G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca NY, 1979.
4. M. Rubinstein and R. H. Colby, *Polymer Physics*, Oxford University Press, Oxford, 2003.
5. A. R. Khokhlov A. Yu Grosberg, *Statistical Physics of Macromolecules*, AIP Press, New York, 1994.
6. K. Binder, ", in: Monte Carlo and Molecular Dynamics Simulations in Polymer Science, K. Binder, (Ed.), p. 356. Oxford University Press, New York, 1995.
7. M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*, Clarendon, Oxford, 1986.
8. P. E. Rouse, **21**, 1272, 1953.
9. F. Bueche, **22**, 603, 1954.
10. W. Paul, G. D. Smith, D. Y. Yoon, B. Farago, S. Rathgeber, A. Zirkel, L. Willner, and D. Richter, *Chain Motion in an Unentangled Polyethylene Melt: A Critical Test of*

the Rouse Model by MD Simulations and Neutron Spin Echo Spectroscopy, **80**, 2346, 1998.
11. H. Tao, T. P. Lodge, and E. D. von Meerwall, **33**, 1747, 2000.
12. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *Comparison of simple potential functions for simulating liquid water.*, J. Chem. Phys., **79**, 926–935, 1983.
13. C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6*, J. Comp. Chem., **25**, no. 13, 1656 – 1676, 2004.
14. A. D. MacKerell, *Empirical force fields for biological macromolecules: Overview and issues*, J. Comp. Chem., **25**, no. 13, 1584 – 1604, 2004.
15. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, *Comparison of multiple amber force fields and development of improved protein backbone parameters*, Proteis, **65**, no. 3, 712 – 725, 2006.
16. B. Hess and N. F. A. van der Vegt, *Hydration thermodynamic properties of amino acid analogues: A systematic comparison of biomolecular force fields and water models*, J. Phys. Chem. B, **110**, no. 35, 17616 – 17626, 2006.
17. W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glattli, P. H. Hunenberger, M. A. Kastenholz, C. Ostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, and H. B. Yu, *Biomolecular modeling: Goals, problems, perspectives*, Angew. Chem. Intl. Ed., **45**, no. 25, 4064 – 4092, 2006.
18. M.-E. Lee and N. F. A. van der Vegt, *Molecular thermodynamics of methane solvation in ter-butanol-water mixtures*, J. Chem. Theory Comput., **3**, no. 1, 194 – 200, 2007.
19. K. Kremer and K. Binder, Comp. Phys. Rept., **7**, 259, 1988.
20. K. Binder and W. Paul, Macromolecules, **41**, 4537, 2008.
21. G. S. Grest and K. Kremer, Phys. Rev. A, **33**, 3628, 1986.
22. K. Kremer and G. S. Grest, J. Chem. Phys., **92**, 5057, 1990.
23. R. Auhl, R. Everaers, G. S. Grest, K. Kremer, and S. J. Plimpton, J. Chem. Phys., **119**, 12718, 2003.
24. T. Soddemann, B. Dünweg, and K. Kremer, Phys. Rev. E, **68**, 046702, 2003.
25. C. F. Abrams and K. Kremer, *Combined coarse-grained and atomistic simulation of liquid bisphenol A-polycarbonate: Liquid packing and intramolecular structure*, Macromolecules, **36**, no. 1, 260 – 267, 2003.
26. V. G. Mavrantzas, T. D. Boone, E. Zervopoulou, and D. N. Theodorou, Macromolecules, **32**, 5072, 1999.
27. A. Uhlherr, S. J. Leak, N. E. Adam, P. E. Nyberg, M. Doastakis, V. G. Mavrantzas, and D. N. Theordorou, *Large scale atomistic polymer simulations using Monte Carlo methods for parallel vector processors*, Comp. Phys. Comm., **144**, 1, 2002.
28. K. Kremer and G. S. Grest, in: Monte Carlo and Molecular Dynamics Simulations in Polymer Science, K. Binder, (Ed.), p. 194. Oxford University Press, New York, 1995.
29. K. Kremer, "Polymer dynamics: Long time simulations and topological constraints.", in: Simulations in Condensed Matter: From Materials to Chemical Biology, Ferrario M., G. Cicotti, and K. Binder, (Eds.), vol. 704 of *Lect. Notes. Phys.* Springer, 2006.
30. T. C. B. Mc Leish, Adv. Phys., **51**, 1379, 2002.
31. S. F. Edwards, Proc. Phys. Soc., **92**, 9, 1967.

32. S. F. Edwards, Proc. Phys. Soc., **91**, 513, 1967.
33. R. Everaers, S. K. Sukumaran, G. S. Grest, C. Svaneborg, A. Sivasubramanian, and K. Kremer, *Rheology and microscopic topology of entangled polymeric liquids*, Science, **303**, 823–826, FEB 6 2004.
34. S. K. Sukumaran, G. S. Grest, K. Kremer, and R. Everaers, *Identifying the primitive path mesh in entangled polymer liquids*, J. Pol. Sci. B, **43**, 917–933, 2005.
35. N. Uchidaa, G. S. Grest, and R. Everaers, *Viscoelasticity and primitive path analysis of entangled polymer liquids: From F-actin to polyethylene*, J. Chem. Phys., **128**, 2008.
36. L. J. Fetters, D. J. Lohse, S. T. Milner, and W. W. Graessley, Macromolecules, **32**, 6847, 1999.
37. M. Pütz, K. Kremer, and G. S. Grest, **49**, 735, 2000.
38. S. Leon, L. Delle Site, N. van der Vegt, and K. Kremer, Macromolecules, **38**, 8078, 2005.
39. L. J. Fetters, D. J. Lohse, and W. W. Graessley, J. Poly. Sci. B: Pol. Phys., **37**, 1023, 1999.
40. S. J. Marrink, D. P. Tieleman, and A. E. Mark, *Molecular dynamics simulation of the kinetics of spontaneous micelle formation.*, J. Phys. Chem. B, **104**, 12165–12173, 2000.
41. C. F. Lopez, S. O. Nielsen, P. B. Moore, and M. L. Klein, *Understanding nature's design for a nanosyringe.*, Proc. Natl. Acad. Sci. USA, **101**, 4431–4434, 2004.
42. M. Muller, K. Katsov, and M. Schick, *Biological and synthetic membranes: What can be learned from a coarse-grained description?*, Physics Reports, **434**, no. 5-6, 113 – 176, 2006.
43. G. S. Ayton, W. G. Noid, and G. A. Voth, *Multiscale modeling of biomolecular systems: in serial and in parallel*, Curr. Opin. Struct. Biol., **17**, no. 2, 192 – 198, 2007.
44. P. L. Freddolino, A. Arkhipov, A. Y. Shih, Y. Yin, Z. Chen, and K. Schulten, "Application of residue-based and shape-based coarse graining to biomolecular simulations.", in: Coarse-Graining of Condensed Phase and Biomolecular Systems, G. A. Voth, (Ed.). Chapman and Hall/CRC Press, Taylor and Francis Group, 2008.
45. N. F. A. van der Vegt, C. Peter, and K. Kremer, "Structure-based coarse- and fine-graining in soft matter simulations", in: Coarse-Graining of Condensed Phase and Biomolecular Systems, G. A. Voth, (Ed.). Chapman and Hall/CRC Press, Taylor and Francis Group, 2008.
46. L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink, *The MARTINI Coarse-Grained Force Field: Extension to Proteins*, J. Chem. Theor. Comput., **4**, no. 5, 819–834, 2008.
47. A. P. Lyubartsev and A. Laaksonen, *Calculation of effective interaction potentials from radial-distribution functions - a reverse Monte-Carlo approach*, Phys. Rev. E, **52**, no. 4, 3730 – 3737, 1995.
48. F. Müller-Plathe, *Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back*, ChemPhysChem, **3**, no. 9, 754 – 769, 2002.
49. C. Peter, L. Delle Site, and K. Kremer, *Classical simulations from the atomistic to the mesoscale: coarse graining an azobenzene liquid crystal*, Soft Matter, **4**, 859–869, 2008.

50. V. A. Harmandaris, N. P. Adhikari, N. F. A. van der Vegt, and K. Kremer, *Hierarchical modeling of polystyrene: From atomistic to coarse-grained simulations*, Macromolecules, **39**, no. 19, 6708 – 6719, 2006.
51. V. A. Harmandaris, D. Reith, N. F. A. van der Vegt, and K. Kremer, *Comparison between Coarse-Graining Models for Polymer Systems: Two Mapping Schemes for Polystyrene*, Macromol. Chem. Phys., **208**, 2109 – 2120, 2007.
52. A. Villa, C. Peter, and N. F. A. van der Vegt, *Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation*, Phys. Chem. Chem. Phys., 2009.
DOI: 10.1039/b818144f
53. A. Villa, N. F. A. van der Vegt, and C. Peter, *Self-assembling dipeptides: including solvent degrees of freedom in a coarse-grained model*, Phys. Chem. Chem. Phys., 2009.
DOI: 10.1039/b818146m
54. C. F. Abrams and K. Kremer, *Effects of excluded volume and bond length on the dynamics of dense bead-spring polymer melts*, J. Chem. Phys., **116**, no. 7, 3162 – 3165, 2002.
55. G. Milano, S. Goudeau, and F. Muller-Plathe, *Multicentered Gaussian-based potentials for coarse-grained polymer simulations: Linking atomistic and mesoscopic scales*, J. Polym. Sci. Pol. Phys., **43**, no. 8, 871 – 885, 2005.
56. J. D. Weeks, D. Chandler, and H. C. Andersen, *Role of repulsive forces in determining equilibrium structure of simple liquids*, J. Chem. Phys., **54**, no. 12, 5237 – 5247, 1971.
57. H. Meyer, O. Biermann, R. Faller, D. Reith, and F. Muller-Plathe, *Coarse graining of nonbonded inter-particle potentials using automatic simplex optimization to fit structural properties*, J. Chem. Phys., **113**, no. 15, 6264 – 6275, 2000.
58. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C. The art of scientific computing.*, Cambridge University Press, 2nd. edition, 1992.
59. D. Reith, M. Putz, and F. Muller-Plathe, *Deriving effective mesoscale potentials from atomistic simulations*, J. Comp. Chem., **24**, no. 13, 1624 – 1636, 2003.
60. J. D. McCoy and J. G. Curro, *Mapping of Explicit Atom onto United Atom Potentials*, Macromolecules, **31**, 9362–9368, 1998.
61. M. E. Johnson, T. Head-Gordon, and A. A. Louis, *Representability problems for coarse-grained water potentials*, J. Chem. Phys., **126**, no. 14, 144509, 2007.
62. W. Tschöp, K. Kremer, J. Batoulis, T. Burger, and O. Hahn, *Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates*, Acta Polym., **49**, no. 2-3, 61 – 74, 1998.
63. B. Hess, S. Leon, N. van der Vegt, and K. Kremer, *Long time atomistic polymer trajectories from coarse grained simulations: bisphenol-A polycarbonate*, Soft Matter, **2**, no. 5, 409 – 414, 2006.
64. B. Hess, C. Peter, T. Ozal, and N. F. A. van der Vegt, *Fast growth thermodynamic integration: solubilities of additive molecules in polymer microstructures*, Macromolecules, **41**, 2283–2289, 2008.
65. V. A. Harmandaris and K. Kremer, in preparation.
66. V.A. Harmandaris and K. Kremer, *Dynamics of Polystyrene Melts through Hierarchical Multiscale Simulations*, Macromolecules, **42**, no. 3, 791–802, 2009.

Adaptive Resolution Schemes

Christoph Junghans, Matej Praprotnik[†], and Luigi Delle Site

Max Planck Institute for Polymer Research
Ackermannweg 10, 55128 Mainz, Germany
E-mail: dellsite@mpip-mainz.mpg.de

The Adaptive Resolution Scheme (AdResS) is a simulation method, which allows to perform Molecular Dynamics (MD) simulations treating different regions with different molecular resolutions. The different scales are coupled within a unified approach by changing the number of degrees of freedom on the fly and preserving the free exchange of particles between regions of different resolution. Here we describe the basic physical principles of the algorithm and illustrate some of its relevant applications.

1 Introduction

Multiscale techniques are becoming standard procedures to study systems in condensed matter, chemistry and material science via simulation. The fast progress of computer technology and the concurrent development of novel powerful simulation methods has strongly contributed to this expansion. This led to the result that detailed sequential studies (modeling) from the electronic scale to the mesoscopic and even continuum are nowadays routinely performed (see e.g.[1–8]). However, sequential approaches still do not couple scales in a direct way. Their central idea is to employ results from one scale to build simplified models in a physically consistent fashion, keeping the modeling approximations as much as possible under control; next, in a separate stage, a larger scale is considered. A step beyond these sequential schemes is represented by those approaches where the scale are coupled in a concurrent fashion within a unified computational scheme. Problems as edge dislocation in metals or crack of materials where the local chemistry effects large scale material properties and vice versa, are typical examples where the idea of concurrent scale methods has been applied. In this case quantum based methods are interfaced with classical atomistic and continuum approaches within a single computational scheme[9–11]. A further example is the Quantum Mechanics/Molecular Mechanics (QM/MM) scheme[12]; mainly used for soft matter systems it is based on the idea that a fixed subsystem is described with a quantum resolution while the remainder of the system is treated at classical atomistic level. A typical example of application of the QM/MM method is the study of the solvation process of large molecules; for this specific example the interesting chemistry happens locally within the region defined by few solvation shells and thus it is treated at a quantum level while the statistical/thermodynamical effect of the fluctuating environment (solvent) far from the molecules is treated in a rather efficient way at classical level. In the same fashion there are several more examples (see e.g. Refs.[13,14]). All of these methods, although computationally robust, are characterized by a non-trivial conceptual limitation, i.e. the region of high resolution is fixed and thus the exchange of particles among the different regions is not allowed. While this may not be a crucial point for hard matter, is certainly a strong limitation for soft matter, i.e. complex fluids, since relevant density fluctuations are arbitrarily

[†]On leave from the National Institute of Chemistry, Hajdrihova 19, SI-1001 Ljubljana, Slovenia

suppressed. The natural step forward to overcome this problem is the design of adaptive resolution methods which indeed allow for the exchange of particles among regions of different resolution. In general, in such a scheme a molecule moving from a high resolution region to a lower one, would gradually lose some degrees of freedom (DOFs) until the lower resolution is reached and yet the statistical equilibrium among the two different regions is kept at any instant. Recently some schemes based on this idea, for classical MD, have been presented in literature[15–19]. They are based on different conceptual approaches regarding the way the scales are coupled and the way the equilibrium of the overall system is assured. For the quantum-classical case there are instead several conceptual problems to be solved before a proper scheme can be designed; this is briefly discussed in the next section.

2 Classical and Quantum Schemes

As stated before, many problems in condensed matter, material science and chemistry are multiscale in nature, meaning that the interplay between different scales plays the fundamental role for the understanding of relevant properties as reported in the examples above. An exhaustive description of the related physical phenomena requires in principle the simultaneous treatment of all the scales involved. This is a prohibitive task not only because of the computational resources but above all because the large amount of produced data would mostly contain information not essential to the problem analyzed and may overshadow the underlying fundamental physics or chemistry of the system. A solution to this problem is that of treating in a simulation only those DOFs, which are strictly required by the problem. In this lecture, in particular, we will illustrate the basic physical principles of the Adaptive Resolution Scheme (AdResS) method, where the all-atom classical MD technique will be combined with the coarse grained MD one (for a general discussion about coarse graining see the contribution of C. Peter and K. Kremer), and briefly discuss the difference with other methods. In the AdResS method the combination of all-atom classical MD and coarse grained MD leads to a hybrid scheme where the molecule can adapt its resolution, passing from an all-atom to a coarse grained representation when going from the high resolution region to the lower one (and vice versa), and thus changing in a continuous manner the number of DOFs on the fly. In this way the limitation of the all-atom approach in bridging the gap between a wide range of length and time scales is overcome by the fact that only a limited region is treated with atomistic DOFs (where high resolution is necessary) while the remaining part of the system is treated in the coarse grained representation and thus loses the atomistic (chemical) details but retains those DOFs relevant to the particular property under investigation. This means that one can reach much longer length and time scales and yet retain high resolution where strictly required. In principle the same concept may be applied for quantum-classical hybrid adaptive schemes. Here for quantum is meant that the nuclei are classical objects but their interaction is determined by the surrounding electrons obeying the Schrödinger equation. In this case, however the level of complexity is by far much higher than the hybrid all-atom/coarse grained case. In fact it involves not only a change of molecular representation but also of the physical principles governing the properties of the system. One of the major obstacles is that of dealing with a quantum subsystem where the number of electrons changes continuously in time, that is the wavefunction normalization varies in time. In this case one deals with

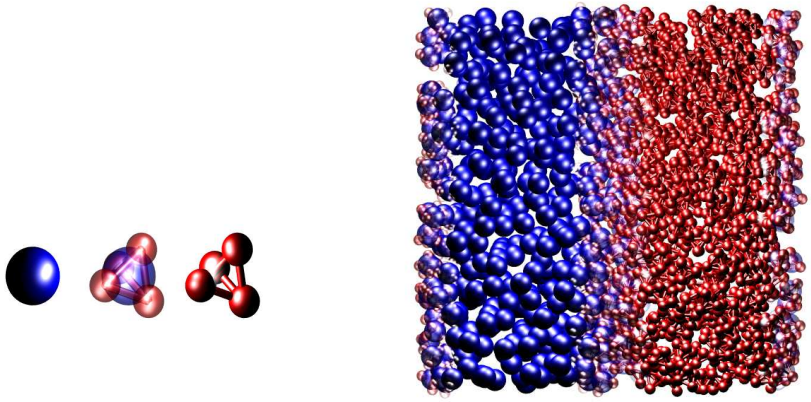
a different Schrödinger problem at each step unless one introduces some artificial creation and annihilation terms in the Hamiltonian in order to allow a continuous fluctuation of the electron number in a consistent way. Although not trivial, this may still be feasible but the physics of the system could be dramatically modified by the presence of such technical artifacts. One should be also careful in not confusing a proper adaptive scheme, where the DOFs (classical and quantum) change continuously on the fly, with the straightforward approach of running a QM/MM-like simulation and at **each step** modify the size of the quantum region. In this case one has a brute force, by-hand adaptivity which does not allow the system to properly relax both the classical and quantum DOFs. A possible solution to the problems above may be that of treating the electron in a statistical way within a macrocanonical ensemble where their number is allowed to fluctuate, along the same line of thinking of Alavi's theory in the Free Energy MD scheme[20], or by mapping the quantum problem of the subsystem into a classical one in a path integral quantum mechanical fashion (see e.g.[21]) so that the idea of adaptivity can be applied between two (effective) classical descriptions. A possible further approach may be along the lines of coupled quantum-classical MD schemes where the classical bath provides the average environment for a quantum evolution of a subsystem via the use of Wigner transformations[22]. However at this stage these are only speculations and up to now no proper quantum-classical procedures where the adaptivity occurs in a continuum smooth way have been proposed.

3 AdResS: General Idea

The driving idea of the AdResS is to develop a scheme where the interchange between the atomistic and coarse level of description is achieved on the fly by changing the molecular DOFs. In order to develop this idea a test model for the molecule has been built. Fig. 1 gives a pictorial representation of the tetrahedral molecule used and its corresponding spherical coarse grained representation, derived in a way that it reproduces chosen all-atom properties. The tetrahedral molecule consists of four atoms kept together by a spring-like potential with a Lennard-Jones intermolecular potential; specific technical details of the model as well as of the coarse grained procedure for the spherical representation are reported in Appendix. As Fig. 1 shows, the atomistic molecule when passing to the coarse grained region, slowly loses its vibrational and rotational DOFs, passing through different stages of hybrid atomistic/coarse grained representation and finally reducing its representation to a sphere whose DOFs are solely the translational ones of the center of mass with a proper excluded volume. A crucial point to keep in mind is that the different resolutions do not mean that the molecules are of different **physical** species. The basic underlying physics is in principle the same in all region and thus the process of exchange has to happen in condition of thermodynamical and statistical equilibrium which means pressure balance $P^{\text{atom}} = P^{\text{cg}}$, thermal equilibrium $T^{\text{atom}} = T^{\text{cg}}$, and no net molecular flux $\rho^{\text{atom}} = \rho^{\text{cg}}$. This conditions must be preserved by the numerical scheme and thus represent the conceptual basis of the method[17], next the effective dynamical coupling between the scales must be specified; this is reported in the next section.

3.1 Scale coupling

Once the effective potential is derived on the basis of the reference all-atom system (see Appendix 8) then the atomistic and the coarse grained scales are coupled via a position



(a) On-the-fly interchange of a molecule (b) Snapshot of the simulation box

Figure 1. (a) The on-the-fly interchange between the atomic and coarse grained levels of description. The middle hybrid molecule is a linear combination of a fully atomistic tetrahedral molecule with an additional center-of-mass particle representing the coarse grained molecule. (b) Snapshot of the hybrid atomistic/mesoscopic model at $\rho^* = 0.1$ and $T^* = 1$ (LJ units). The red molecules are the explicit atomistically resolved tetrahedral molecules, while the blue molecules are the corresponding one-particle coarse grained molecules. (Figure was taken from Ref.[15])

dependent interpolation formula on the atomistic and coarse grained force[15,16]:

$$\mathbf{F}_{\alpha\beta} = w(X_\alpha)w(X_\beta)\mathbf{F}^{\text{atom}}_{\alpha\beta} + [1 - w(X_\alpha)w(X_\beta)]\mathbf{F}^{\text{cg}}_{\alpha\beta} \tag{1}$$

where α and β labels two distinct molecules, $\mathbf{F}^{\text{atom}}_{\alpha\beta}$ is derived from the atomistic potential where each atom of molecule α interacts with each atom of molecule β, and $\mathbf{F}^{\text{cg}}_{\alpha\beta}$ is obtained from the effective (coarse grained) pair potential between the centers of masses of the coarse grained molecules. In the region where a smooth transition from one resolution to another takes place, a continuous monotonic "switching" function $w(x)$ is defined as in Fig. 2 (where X_α, X_β are the x-coordinates of the centers of mass of the molecules α and β). A simple way to think about the function $w(x)$ is the following: $w(x)$ is equal to one in the atomistic region and thus the switchable DOFs are fully counted, while $w(x)$ is zero in the coarse grained region and thus the switchable DOFs are turned off, while in between takes values between zero and one and thus provides (continuous) hybrid representations of such DOFs (i.e. they count only in part). In general, Eq. 1, allows for a smooth transition from atomistic to coarse grained trajectories without perturbing the evolution of the system in a significant way. More specifically the formula of Eq. 1 works in such a way that when a molecule passes from the atomistic to the coarse grained region, the molecular vibrations and rotations become less relevant until they vanish so that $w(x)$ smoothly "*freezes*" the dynamical evolution of these DOFs and their contributions to the interaction with the other molecules. Vice versa, when the molecules goes from the coarse grained region to the atomistic one, $w(x)$ smoothly "*reactivates*" their dynamics and their contributions to the intermolecular interactions. In the case of tetrahedral molecules, being characterized by pair interactions, we have that all the molecules interacting with coarse grained molecules interact as coarse grained molecules independently of the region

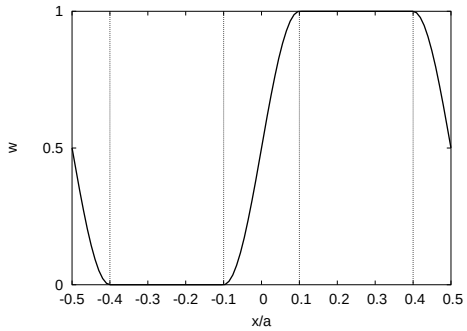


Figure 2. The weighting function $w(x) \in [0, 1]$. The values $w = 1$ and $w = 0$ correspond to the atomistic and coarse grained regions of the hybrid atomistic/mesoscopic system with the box length a, respectively, while the values $0 < w < 1$ correspond to the interface layer. The vertical lines denote the boundaries of the interface layers. (Figure was taken from Ref.[15])

where they are (the coarse grained molecule does not have any atomistic detail, thus the other molecule can interact with this molecule only via the center of mass), two atomistic interact as atomistic, while for the other couplings, the interactions are governed by the $w(X_\alpha)w(X_\beta)$ combination. A very important point of Eq. 1 is that, by construction, the Newton's Third Law is preserved. The diffusion of molecules between regions with different resolution must not be perturbed by the resolution change. Thus the conservation of the linear momentum dictated by the Newton's Third Law is crucial in adaptive resolution MD simulations.

3.2 Thermodynamical equilibrium

Eq. 1 cannot be derived from a potential and thus a scheme based on it, would not have an energy to conserve. The natural subsequent question is how to then control the thermodynamic equilibrium. The conceptual problem for an adaptive scheme is that the free energy density is formally not uniform since the number of DOFs varies in space, however being the system uniform by construction (and being the **underlying physical** nature of the molecules the same everywhere), this would be only an artifact of the formalism used. This non uniformity leads to a non-physical preferential direction of the molecular flux. In fact, as numerical tests show, there is a preferential tendency of the atomistic molecules to migrate into the coarse grained region and change resolution in order to lower the free energy of the system (the free energy is an extensive quantity, that is proportional to the number of DOFs). A simple qualitative way to picture this diode-like aspect is the following: when a molecule goes from an atomistic to a coarse grained region it loses vibrational and rotational DOFs and thus in its interactions with the neighboring (coarse grained) molecules it must accommodate only its excluded volume (i.e. find space). This becomes more complicated if a coarse grained molecules moves into an atomistic region, in this case the molecule acquires rotational and vibrational DOFs and tries to enter into a region where other molecules are already locally in equilibrium. This means that in order to enter this region, the molecules should accommodate both rotational and vibrational DOFs

according to the neighboring environment. Most likely the molecule would enter with vibrational and rotational motions which does not fit the local environment and this would lead to a perturbation of the local equilibrium. This means that for such a molecule the **way back** to the coarse grained region is more convenient, and thus this free energy barrier works as a closed door (probabilistically) for the coarse grained molecules and opened door for the atomistic ones so that a preferential molecular flux from the atomistic to the coarse grained region is produced. In thermodynamic terms, as an artifact of the method, the different regions are characterized by a different chemical potential, however, since this aspect does not stem from the physics of the system but only from the formalism, we have to amend for this thermodynamical unbalance. This means that the use of Eq. 1 alone cannot assure thermodynamical equilibrium and further formal relations, linking the variables of the problem, should be determined in order to obtain equilibrium. This can be obtained, as shown in the next sections, by analyzing the meaning of the process of varying resolution in statistical and thermodynamical terms.

4 Theoretical Principles of Thermodynamical Equilibrium in AdResS

In this section we analyze the idea of describing thermodynamical equilibrium for a system where, formally, the number of DOFs is space dependent and yet the molecular properties are uniform in space.

4.1 The principle of geometrically induced phase transition

The space dependent change of resolution can be seen, to have some similarities to a physical phase transition, as a fictitious geometrically induced phase transition. In simple words, the concept of latent heat is similar to that of a molecule which, for example, goes from the liquid to the gas phase and in doing so needs a certain energy (latent heat) to activate those vibrational states that makes the molecules free from the tight bonding of the liquid state. In the same way, a molecule in the adaptive scheme that passes from a coarse grained to an atomistic resolution, needs a latent heat to formally (re)activate the vibrational and rotational DOFs and to reach equilibrium with the atomistic surrounding. Vice versa the heat is released when the molecule goes from gas to liquid and so the bond to the other molecules becomes tighter, in the same way in the adaptive scheme, the molecule passing from atomistic to coarse grained, formally releases DOFs and thus automatically the associated heat. This concept can be formalized as: $\mu^{\text{atom}} = \mu^{\text{cg}} + \phi$, where μ^{cg} is the chemical potential calculated with the coarse grained representation, μ^{atom} that of the atomistic one, and ϕ is the latent heat[23, 17]. Possible procedures for a formal derivation of an analytic or numerical form of ϕ and how to use it in the AdResS scheme is still a matter of discussion and subject of work in progress and will be briefly discussed later on. For the time being, a simpler and practical solution is used, that is the system is coupled to a thermostat (see Appendix 8) which automatically, as a function of the position in space, provides (or removes) the required latent heat assuring stability to the algorithm and equilibrium to the system. The coupling of the system to a thermostat leads to the natural question of how to define the temperature in the region of transition where the number of DOFs is space dependent.

4.2 Temperature in the transition region

In the atomistic and coarse grained region the temperature can be defined without a problem employing the equipartition theorem: $T^{\text{atom/cg}} = 2\frac{<K^{\text{atom/cg}}>}{n^{\text{atom/cg}}}$, where $< K^{\text{atom/cg}} >$ is the average kinetic energy of the atomistic/coarse grained region and $n^{\text{atom/cg}}$ is the total average number of DOFs. In the atomistic/coarse grained region, such a quantity is a well defined number, however it is not so in the transition region where $n^{\text{trans}} = n(x)$. The question arising is how to define T^{trans} and above all what $< K^{\text{trans}} >$ means. To address this question we make the following observations: the switching procedure implies that a DOF, in calculating average statistical quantities, **fully counts** in the atomistic region, which formally means that an integral over its related phase space is performed ($\int ....\text{d}q$; q being a generic switchable DOF). On the other hand in the coarse grained region, q is not relevant to the properties of the system and thus it **does not count at all**, that is no integration over its related phase space is required. In the transition region the situation is something in between and thus by switching on/off the DOF q we effectively change the dimensionality (between zero and one) of its related phase space, that is of its domain of integration. In simple words q in the transition region contributes to statistical averages with a weight. The mathematical tool which allows to formalize this idea is provided by the technique of the fractional calculus, where for a fixed resolution w the infinitesimal volume element is defined as[26]:

$$\text{d}V_w = \text{d}^w q\, \Gamma(w/2)/2\pi^{w/2}\Gamma(w) = |q|^{w-1}\text{d}q/\Gamma(w) = d\text{q}^w/w\Gamma(w) \tag{2}$$

with $\Gamma(w)$ the well-known Γ function. Employing such a formalism to calculate the average energy for quadratic DOFs one obtains:

$$< K_q >_w = \frac{\int_0^\infty e^{-\beta q^2} q^{w+1}\text{d}q}{\int_0^\infty e^{-\beta q^2} q^{w-1}\text{d}q}. \tag{3}$$

The solution of Eq. 3 is found to be[26]:

$$< K_q >_w = \frac{w}{2}\beta^{-1}. \tag{4}$$

This is nothing else than the analog of the equipartition theorem for non integer DOFs. Here $< K_q >_w$ is the average kinetic energy of the switchable DOF q for the fixed resolution w. One can then think to use w as a continuous parameter and thus obtaining the definition of kinetic energy for the switchable DOFs in the transition region. A further point needs to be explained, that is, we have implicitly used a Hamiltonian to perform the ensemble average and this would contradict the statement of the previous section about the non existence of an energy within the coupling scheme used. To clarify this aspect we have to say that the coupling formula on the forces is not directly related to the derivation of the statistical average performed here. Here we have interpreted the process of changing resolution as the process of partially counting a DOF contribution into the statistical determination of an observable, under the hypothesis that the underlying Hamiltonian is the same all over the system. This is justified by the fact that the underlying physics is in principle the same all over the system but the formal representation and thus the analysis of the DOFs of interest and their contributions differs. This in practical terms means that the derivation of the temperature and the principle of coupling of forces via spatial interpolation are two aspects of the same process but one cannot formally derive both from

a single general principle so that the connection between them, at this stage, must be intended as only qualitative. However, we will use the numerical tool of simulation where both Eq. 1 and Eq. 4 are employed in connection to each other to prove that they are numerically consistent. At this point the obvious question arises about why to choose the approach based on the interpolation of the forces and not to choose the more natural one based on the smooth interpolation of the potential. This problem is treated in the next section.

5 Coupling the Different Regimes via a Potential Approach

The coupling scheme analog of Eq. 1 using potentials instead of forces would be:

$$U_{\alpha\beta} = w(x_\alpha)w(x_\beta)U^{\text{atom}}_{\alpha\beta} + [1 - w(x_\alpha)w(x_\beta)]U^{\text{cg}}_{\alpha\beta}. \tag{5}$$

This approach leads to a series of problem whose solution is not trivial. In particular if one derives the forces from Eq. 5 obtains an extra term, which here we will name **drift force**, of the following form:

$$\mathbf{F}^{\text{drift}} = U^{\text{atom}}\frac{\partial w}{\partial x} + U^{\text{cg}}\frac{\partial w}{\partial x} \tag{6}$$

There are two options at this point, one accepts this force as a result of a definition of a new force field in Eq. 5, or one tries to remove it by a specific choice of $w(x)$ or by modifying $U_{\alpha\beta}$ in Eq. 5. In the first case one has to be aware that, because the derivative of $w(x)$ enters into the equations of motion, the evolution of the system becomes highly sensitive to the choice of the form of $w(x)$. This means that different functions $w(x)$ may lead to complete different results, and being the choice of $w(x)$ made on empirical basis, the dynamic becomes arbitrary and thus, most likely, unphysical. The limitation above applies in principle to the approach proposed by Heyden and Truhlar[19], where the scales are coupled by an interpolation of Lagrangians via a space dependent function. Moreover, the force obtained from Eq. 5 does not preserve the third Newton's law[23,26].

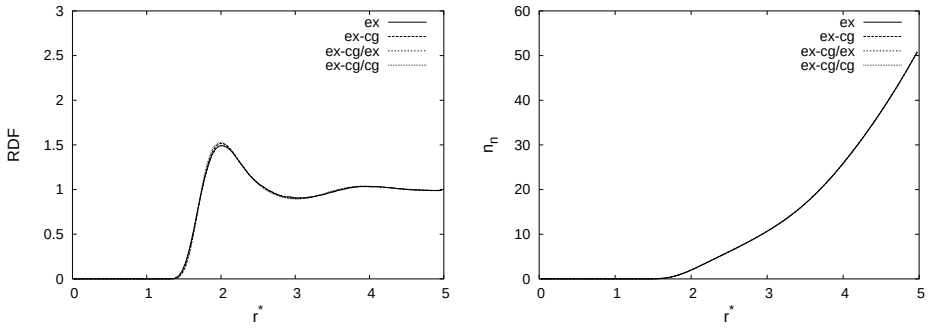
Instead if one tries to follow the second possibility, that is removing $\mathbf{F}_{drift}$, one encounters heavy mathematical difficulties[24,25] since the condition $\mathbf{F}_{drift} = 0$ leads to a system of partial differential equations of first order:

$$\begin{aligned} U^{\text{cg}}\frac{\partial f(X_\alpha, X_\beta)}{\partial X_\alpha} + U_{atom}\frac{\partial g(X_\alpha, X_\beta)}{\partial X_\alpha} = 0 \\ U^{\text{cg}}\frac{\partial f(X_\alpha, X_\beta)}{\partial X_\beta} + U_{atom}\frac{\partial g(X_\alpha, X_\beta)}{\partial X_\beta} = 0. \end{aligned} \tag{7}$$

Here $f(x)$ and $g(x)$ are the most general switching functions one can think of. For the system of Eqs. 7 each equation is characterized by two boundary conditions, thus the system is **overdetermined** and thus in general a solution **does not exist**. This is valid also if one tries to generalize Eq. 5 as:

$$U^{\text{coupling}} = f(X_\alpha, X_\beta)U^{\text{cg}} + g(X_\alpha, X_\beta)U^{\text{atom}} + \Phi. \tag{8}$$

The extra potential Φ does not improve the situation because in this case the overdetermination is shifted from f and g to Φ. These sort of problems, in principle, occur for the conserving energy method proposed by Ensing *et al.*[18], where the difference between the



(a) Radial distribution functions (b) Average numbers of neighbors

Figure 3. (a) Center-of-mass radial distribution functions for all molecules in the box of the all-atom ex and hybrid atomistic/mesoscopic ex-cg systems at $\rho^* = 0.1$ and $T^* = 1$. Shown are also the corresponding center-of-mass radial distribution functions for only the explicit molecules from the explicit region ex-cg/ex and for only the coarse grained molecules from the coarse grained region ex-cg/cg . The width of the interface layer is $2d^* = 2.5$. (b) The corresponding average numbers of neighbors $n_n(r^*)$ of a given molecule as functions of distance. The different curves are almost indistinguishable. (Figure was taken from Ref.[15])

true (full atomistic) energy of the system and the one of the hybrid scheme is provided during the adaptive run via a book keeping approach while the forces are calculated with a scheme similar to that of AdResS. The problem of the overdetermination reported above in this case would mean that the conserved energy is not consistent with the dynamics of the system. In comparison, the AdResS method has the limitation of not even attempting to define an energy but on the other hand the overall scheme is robust enough to keep the dynamics and the essential thermodynamics under control without the problem of energy conservation. The next step consists of using the information gained so far and apply the principles of the previous section in a numerical experiment to prove the validity of the scheme.

6 Does the Method Work?

In order to prove that such a computational approach with the theoretical framework presented so far is robust enough to perform simulations of chemical and physical systems we have carried on studies for the liquid system of tetrahedral molecules where the results of the AdResS approach are compared with the results obtained with full atomistic simulations. First we have shown that global and local structure can be reproduced. This means, we have determined the center of mass-center of mass radial distribution function for the whole system (global), and for only the atomistic part and only for the coarse grained part (local) and compared it with the results obtained in a full atomistic simulation. This comparison for a medium dense liquid are reported in Fig. 3; the agreement is clearly satisfactory since the various plots are all on top of each other. However the radial distribution function is based on an average over the space, this means that cannot describe possible local and instantaneous fluctuations due to some possible artifact of the method. These latter may not be negligible but, by compensating each other, they would not appear in the

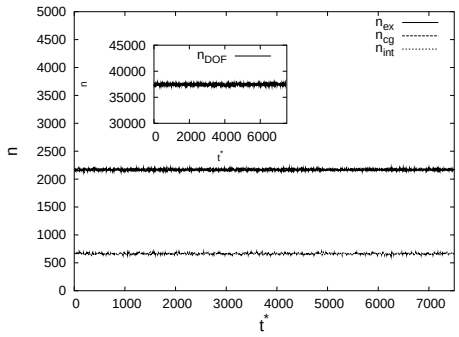


Figure 4. Time evolution of number of molecules in explicit $n_{\rm ex}$, coarse grained $n_{\rm cg}$, and interface regions $n_{\rm int}$ in the hybrid atomistic/mesoscopic model with the 2.5 interface layer width. The time evolution of the number of degrees of freedom in the system $n_{\rm DOF}$ is depicted in the inset. (Figure was taken from Ref.[15])

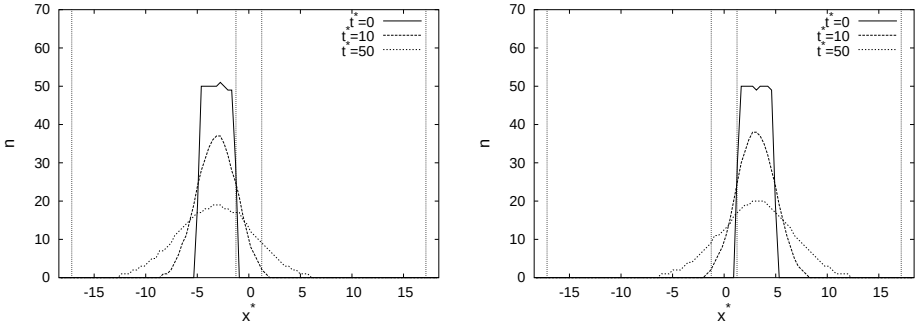


Figure 5. Time evolution of diffusion profiles for the molecules that are initially, at time $t^*=0$, localized at two neighboring slabs of the mid-interface layer with $2d^*=2.5$ (n is the number of these molecules with the center-of-mass position at a given coordinate x^*). The width of the two slabs is $a^*/10$. The vertical lines denote the boundaries of the interface layer. (a) The diffusion profile, averaged over 500 different time origins, at $t^*=0$, $t^*=10$, and $t^*=50$ for the molecules that are initially localized at the slab on the coarse grained side of the interface region. (b) The same as in (a) but for the molecules that are initially localized at the slab on the atomistic side of the interface region. (Figure was taken from Ref.[15])

plot of Fig. 3. In this sense the study above is not sufficient to infer about the validity of the method. Therefore, we also studied the evolution of the number of DOFs as a function of time. This should make us aware of possible non-negligible artificial fluctuations of the system. Fig. 4 shows that the number of DOFs is conserved at any time during the run and thus there is no net flux through the border of the two regions. Again, this study is not sufficient to prove the validity of the scheme, because still one should prove that indeed there is a true exchange of particles from one region to another. In fact it may happen that the equilibrium among the different regions is due to a reflection mechanism without exchange of particles between them. Fig. 5 shows that indeed a sample of molecules from the atomistic region diffuses into the coarse grained one and vice versa a sample from the coarse grained region diffuse into the atomistic one. It is however only a coincidence that

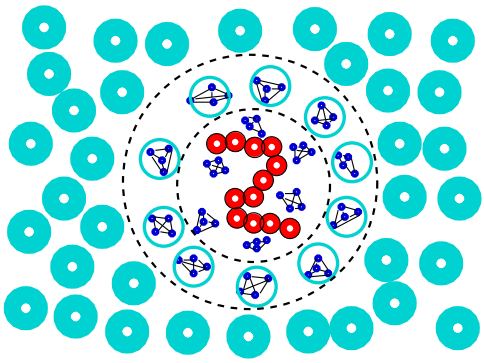


Figure 6. A schematic plot of a solvated generic bead-spring polymer. The solvent is modeled on different levels of detail: solvent molecules within a certain radius from the polymer's center of mass are represented with a high atomistic resolution while a lower mesoscopic resolution is used for the more distant solvent. The high resolution sphere moves with the polymer's center of mass. The polymer beads are represented smaller than the solvent molecules for presentation convenience; for details see text. (Figure was taken from Ref.[27])

this happens in a symmetric way, because the system in question has basically the same diffusion constant in the atomistic and coarse grained representation. In general the profile is not symmetric. To remove this unphysical effect the system is coupled with a position dependent thermostat to match the diffusion constants of the atomistic and coarse grained molecules (see Appendix 8). The data reported in the plots above are for a medium dense liquid[15], but the same satisfactory agreement was found for high density liquid[16].

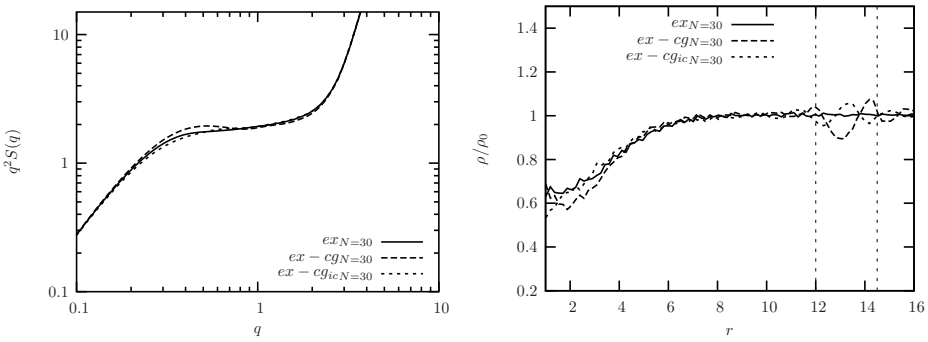
7 Further Applications

7.1 Solvation of a simple polymer in the tetrahedral liquid

An extension of the approach above to a solvation of an ideal bead and spring polymer in tetrahedral liquid was then performed in Ref.[27]. Here the solvation shell is defined as the atomistic region, and outside the solvent is represented with its coarse grained spheres. The solvation shell, centered at the center of mass of the polymer is always large enough that the polymer is contained in it. This region can diffuse around with the polymer and all the molecules entering the solvation area become atomistic and those leaving the region become coarse grained. As for the cubic box before, between the atomistic and the coarse grained regions there is a transition region (see Fig. 6). Two examples of comparison with a full atomistic simulation are reported, these are the calculation of the static form factor (left panel Fig. 7) and the shape of the solvation region as a function of the distance from the center (of the region) in terms of particle density (right panel Fig. 7). These two plots show very good agreement with the full atomistic simulation and thus prove that the method is indeed robust for such a kind of system.

7.2 Liquid water

The first application to a real chemical and physical system is that to liquid water. Several new technical issues arise, the most relevant of which are the presence of the charges and



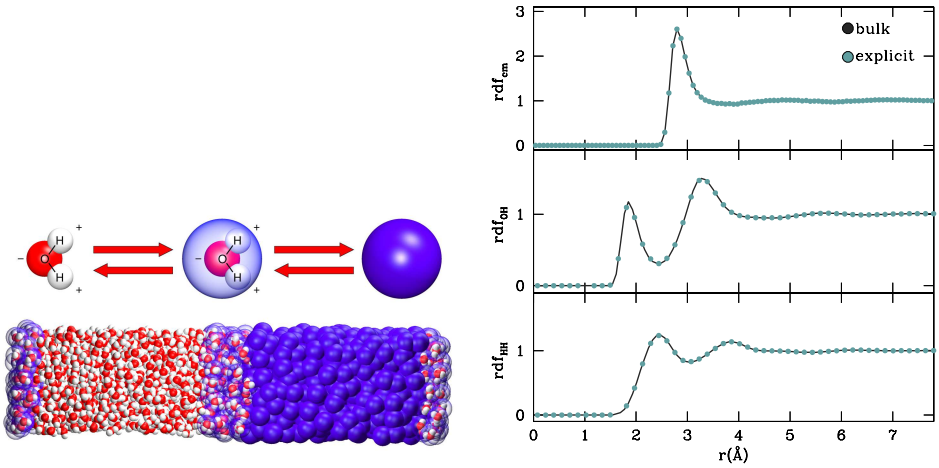
(a) Structure factor (b) Correlation hole

Figure 7. (a) The static structure factor of the polymer with $N = 30$ in the Kratky representation for all three cases studied: the fully explicit, the AdResS scheme with and without the interface pressure correction. (b) The correlation hole for the same systems as in (a) . (Figure was taken from Ref.[27])

the different diffusion coefficients in the atomistic and coarse grained representations (see Fig. 8)[28,29]. These technical problems have been solved and the approach used is reported in the appendix, here we report only the results showing that indeed the adaptive simulation can reproduce in a satisfactory way the results of the full atomistic ones. This is shown in the right panel of Fig. 8(b), where several radial distribution functions calculated in the full atomistic simulation and in the adaptive case (for the atomistic region) are plotted . Moreover, not shown here, results of the study show that the system remains indeed uniform. Several other properties were calculated showing the robustness of such an approach for liquid water and they are reported in Refs.[28,29].

7.3 Triple-scale simulation of molecular liquids

Recently we succeeded in developing a triple scale approach where the atomistic is interfaced with the coarse grained description and the latter with the continuum[30,31]. This multiscale approach was derived by combining two dual-scale schemes: our particle-based AdResS, which links the atomic and mesoscopic scales within a molecular dynamics (MD) simulation framework, and a hybrid flux-exchange based continuum-MD scheme (HybridMD) developed by Delgado-Buscalioni *et al.*[32,33]. The resulting triple-scale model consists of a particle-based micro-mesoscale MD region, which is divided into a central atomistic and a surrounding mesoscopic domain, and a macroscopic region modeled on the hydrodynamic continuum level as schematically presented in Fig. 9 for the example of the tetrahedral liquid. The central idea of the triple-scale method is to gradually increase the resolution as one approaches to the atomistic region, which is the "region of interest". The continuum and MD region exchange information via mass and momentum fluxes, which are conserved across the interface between continuum and MD regions (for details see Refs.[32,33]). The combined approach successfully solves the problem of large molecule insertion in the hybrid particle-continuum simulations of molecular liquids and at the same time extends the applicability of the particle-based adaptive resolution schemes to simulate



(a) Mapping scheme (b) Radial distribution functions

Figure 8. (a) On-the-fly interchange between the all-atom and coarse grained water models. Top: the explicit all-atom water molecule is represented at the right, and the coarse grained molecule at the left. The middle hybrid molecule interpolates between the two (see text). Bottom: a schematic representation of the full system, where a hybrid region connects the explicit and coarse grained levels of description. All the results presented in the paper were obtained by performing N V T simulations using ESPResSo[34] with a Langevin thermostat, with a friction constant $\zeta = 5\text{ps}^{-1}$ and a time step of 0.002ps at $T_{\text{ref}} = 300$K and $\rho = 0.96\text{g/cm}^{-3}$ (the density was obtained from an NPT simulation with $P_{\text{ref}} = 1$bar). Periodic boundary conditions were applied in all directions. The box size is 94.5 Å in the x direction and 22 Å in the y and z directions. The width of the interface layer is 18.9 Å in the x direction. (b)The center-of-mass, OH and HH RDFs for the explicit region in the hybrid system dots, and bulk line systems. (Figures were taken from Refs.[28] and[29])

open systems in the grand-canonical ensemble including hydrodynamic coupling with the outer flow.

8 Work in Progress: Towards an Internally Consistent Theoretical Framework

The AdResS method has been shown to be numerically rather robust, however further developments of the theoretical framework, on which the method is based, would be highly desirable in order to improve the structure and the flexibility of the algorithm. One relevant point regards the concept of latent heat introduced via the theoretical analysis about the meaning of changing resolution in thermodynamical terms. This has been so far implemented numerically by using a thermostat; such an approach is numerically very convenient to stabilize the algorithm and drive the system to equilibrium but at the same time does not permit the detailed control of the physical process occurring while the change of resolution happens. To this aim we are making an effort to formalize the concept of latent heat on the basis of a physical ground by employing first principles of thermodynamics or statistical mechanics. In this way an explicit analytic or semi-analytic description of the latent heat, would allow to avoid the use of a stochastic thermostat and automatically

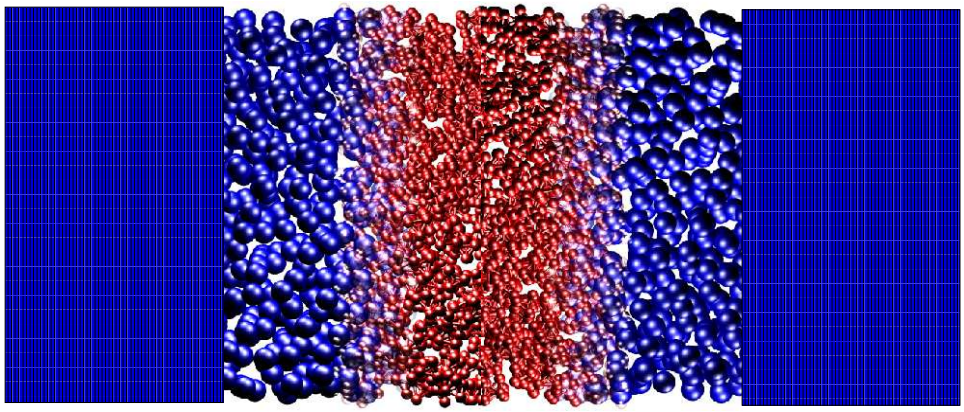


Figure 9. The triple-scale model of the tetrahedral liquid. The molecular particle-based region is embedded in the hydrodynamics continuum described by Navier-Stokes equations (solved by the finite volume method). The molecular region is divided into the central explicit atomistic region with all-atom molecules (red tetrahedral molecules) sandwiched in between two coarse grained domains with the molecules represented on a more coarse grained level of detail (one particle blue molecules). (Figure was taken from Ref.[30])

provide thermodynamic equilibrium. With that the dynamics and the essential thermodynamics can be taken explicitely under control and provide equilibrium despite the fact that still we do not define an energy as in standard simulation schemes. For this purpose we reformulate the problem of the latent heat in terms of an additional thermodynamic force. Such a thermodynamic force is represented by the gradient of a scalar field whose task is that of assuring the balance of the chemical potential in all regions. Such a field can be derived by calculating numerically the chemical potential or the free energy density in the various region of different resolution. Numerical as well as analytic work on this subject is in progress.

Appendix A: Tetrahedral Fluid

In the atomistic representation every molecule of this model fluid consist of 4 atom. All of these have the same mass m_0 and interact via purely repulsive Lennard-Jones potential:

$$U_{\mathrm{LJ}}^{\mathrm{atom}}(r_{\alpha i\beta j}) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r_{\alpha i\beta j}}\right)^{12} - \left(\frac{\sigma}{r_{\alpha i\beta j}}\right)^{6} + \frac{1}{4}\right] & : r_{\alpha i\beta j} \le 2^{1/6}\sigma \\ 0 & : r_{\alpha i\beta j} > 2^{1/6}\sigma \end{cases}, \tag{9}$$

where $r_{\alpha i\beta j}$ is the distance between the ith atom in the αth molecule and the jth atom in the βth molecule, note that we also consider the Lennard-Jones interactions between the atoms of the same molecule. Additionally the atoms in one molecule are bonded by FENE potential

$$U_{\mathrm{FENE}}^{\mathrm{atom}}(r_{\alpha i\alpha j}) = \begin{cases} -\frac{1}{2}kR_0^2 \ln\left[1 - \left(\frac{r_{\alpha i\alpha j}}{R_0}\right)^2\right] & : r_{\alpha i\alpha j} \le R_0 \\ \infty & : r_{\alpha i\alpha j} > R_0 \end{cases}, \tag{10}$$

with a divergence length $R_0 = 1.5\sigma$ and stiffness $k = 30\epsilon/\sigma^2$.

Appendix B: Mapping Scheme/Coarse Grained Potentials

In coarse grained representation we replace a molecule by a bead located at the position of the center of mass of the atomistic molecule. The interaction between the coarse grained beads is determined by the iterative Boltzmann inversion[35] and it is such that the radial distribution function (RDF) of the coarse grained system fits the RDF of the atomistic system. In summary this procedure works as follows (for a detailed presentation see the lecture of C. Peter and K. Kremer). After starting with an initial guess of the pair interaction $V_0(r)$, the interaction of the $(i+1)$th step is given by:

$$V_{i+1}(r) = V_i(r) + \mathrm{k_B}T \ln \left[\frac{g_i(r)}{g^{\text{target}}(r)} \right] , \tag{11}$$

where $g^{\text{target}}(r)$ is the RDF we want to fit, usually given by atomistic simulation and $g_i(r)$ is the RDF of the ith step. Commonly the potential of mean force is used as an initial guess:

$$V_0(r) = -\mathrm{k_B}T \ln g^{\text{target}}(r) \tag{12}$$

Appendix C: Interface Correction

In the switching region the density profile is not uniform, instead it is characterized by some evident fluctuations. Such fluctuations are due to the fact that for hybrid interactions the corresponding effective potential is not the same as the full coarse grained one for matching the structure and the pressure of the atomistic system. Technically this means that we need to derive first an effective potential between hybrid molecules with a fixed weight, which reproduces the RDF and the pressure of the atomistic one, and then, in order to suppress the density fluctuations, use it for an interface correction. Here we report the case $w = 0.5$, however the extension to other weights (and other points) is straightforward. The newly derived effective potential with $w = 0.5$, $V^{\text{ic,0.5}}(R_{\alpha\beta})$ is determined via the iterative Boltzmann procedure (as before, see Appendix 8). Then, one replaces the forces between the coarse grained beads by[16]:

$$\mathbf{F}^{\text{ic}}_{\alpha\beta} = s[w(R_\alpha)w(R_\beta)]\mathbf{F}^{\text{cg}}(R_{\alpha\beta}) + (1 - s[w(R_\alpha)w(R_\beta)])\mathbf{F}^{\text{ic,0.5}}(R_{\alpha\beta}) , \tag{13}$$

where $\mathbf{F}^{\text{ic,0.5}}(R_{\alpha\beta})$ is the force coming from the potential $V^{\text{ic,0.5}}(R_{\alpha\beta})$ and

$$s[x] = 4(\sqrt{x} - 0.5)^2 , \tag{14}$$

is a function $s \in [0, 1]$, which is zero for both weights being 0.5 ($s[(0.5)^2] = 0$) and one for the product of the two weights being 0 or 1 (s[0]=s[1]=0); this means that one has the "exact" force when both molecules have $w = 0.5$. For other weights the force is smoothly interpolated between the corrected and the standard coarse grained force, and thus one obtains an improvement at the interface. In principle if one repeated this procedure for each combinations of $w(R_\alpha)w(R_\beta)$, in the switching region one would have always the exact force. We have noticed that numerically is enough to have a correction for the worst case ($w = 0.5$).

Appendix D: Charged Molecules

Electrostatic interactions are long-ranged and must be calculated over several periodical images of the simulation box. This leads to some problems in the adaptive resolution method because, on one hand, molecules become uncharged in their coarse grained representation and on the other hand the long-ranged character of electrostatic interactions leads to self interaction of all periodical images, for example the interaction of the explicit regions of two image boxes. Additionally, standard approaches like particle mesh Ewald or P3M will always lead to an all-with-all interaction of the molecules, due the involved Fourier transformation, and thus making the switching of the degrees of freedom not possible.

Luckily, in the case of dense and homogeneous fluids (like water) one can use the reaction field approach[36]. The latter assumes that outside a sphere with radius r_{cut} the charges are distributed homogeneously, and thus it makes it possible to replace the interactions outside the sphere with that of a continuum with a dielectric constant ϵ_{rf}. This scheme has been frequently used for liquid water[37], and, in this case, it allows to treat charged molecules in the adaptive resolution method, where one deals with pair interactions:

$$U(r) = \begin{cases} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1}\left[\frac{1}{r} - \frac{Br^2}{2r_c^3} - \frac{2-B}{2r_c}\right] & : r \leq r_c \\ 0 & : r > r_c \end{cases} \tag{15}$$

with $B = 2(1-\epsilon_1-\epsilon_{\text{rf}})/(\epsilon_1+2\epsilon_{\text{rf}})$. The ϵ_{rf} is the dielectric constant outside the cut-off. r_c, which can be estimated from a particle mesh Ewald calculation or determined in a recursive manner.

Appendix E: Thermostat

In general a thermostat is always needed to perform a NVT simulation. Specifically, in the case of the adaptive resolution scheme the thermostat is also needed to compensate for the switch of the interaction between the molecules, since it ensures that the atoms of a molecule have the correct velocity distribution when entering the switching region from the coarse grained side. We use the Langevin idea or stochastic dynamics[38] to ensure the correct ensemble by adding a random and a damping force

$$\dot{\mathbf{p}}_i = -\nabla_i U + \mathbf{F}_i^{\text{D}} + \mathbf{F}_i^{\text{R}} \tag{16}$$

The damping force is Stokes-like force

$$\mathbf{F}_i^{\text{D}} = -\zeta_i/m_i\ \mathbf{p}_i \tag{17}$$

To compensate for this friction one adds a random force

$$\mathbf{F}_i^{\text{R}} = \sigma_i \eta_i(t)\ , \tag{18}$$

where η_i is a noise with zero mean ($<\eta_i(t)>=0$) and certain correlation properties ($<\eta_i(t)\eta_j(t')>=\delta_{ij}\delta(t-t')$. And ζ_i, σ_i are the friction and the noise strength. The corresponding Fokker-Planck operator[39] for the stochastic part of the Langevin equation (Eq. 16) is given by:

$$\mathcal{L}_{\text{SD}} = \sum_i \frac{\partial}{\partial \mathbf{p}_i}\left[\zeta_i \frac{\partial \mathcal{H}}{\partial \mathbf{p}_i} + \sigma_i^2 \frac{\partial}{\partial \mathbf{p}_i}\right]\ , \tag{19}$$

where the sum goes over all particles. By assuming that the equilibrium distribution is a Boltzmann distribution, one has:

$$\mathcal{L}_{\mathrm{SD}} \exp(-\mathcal{H}/\mathrm{k_B}T) = 0 \tag{20}$$

and from that one obtains:

$$\sigma_i^2 = \mathrm{k_B}T\zeta_i \ , \tag{21}$$

which is also known as the Fluctuation-Dissipation theorem (FDT)[39]. At this point we are left with one free parameter to choose, namely the friction strength ζ_i. The drawback of this thermostat is its lack of Galilei invariance and the strong dependence of the dynamics from the friction strength. Therefore, it is in many cases more appropriate to use the Galilei invariant and hydrodynamics conserving DPD thermostat[40], which leaves the dynamic nearly unchanged for wide range of ζ.

Applying the thermostat in AdResS

To obtain the FDT from Eq. 20 a Hamiltonian is needed and, as discussed above, in the AdResS method it is not possible to define a Hamiltonian. For this reason one has to couple the thermostats acting on the explicit and coarse grained molecules. One could make, for example, a linear combination of the thermostat forces (as in Eq. 1 for the deterministic forces). However, this would violate the FDT because the ratio of "random force squared to damping force" would not be conserved (see Eq. 21). Consequently, the temperature would not be correctly defined. Another possibility is to apply the linear scaling to the friction coefficient of the damping force (from the atomistic friction coefficient at the all-atom/transition regime interface to the coarse grained one at the transition/coarse grained boundary) and adjust the noise strength σ to satisfy the FDT[29,30] (see also the next section). The thermostat is then applied to the explicit particles in the atomistic and transition regions and to the center of mass interaction sites in the coarse grained regime. In addition, the explicit atoms of a given molecule, which enters the transition regime from the coarse grained side, are also assigned rotational/vibrational velocities corresponding to atoms of a random molecule from the atomistic region (where we subtract the total linear momentum of the latter molecule). By doing this we ensure that the kinetic energy is distributed among all DOFs according to the equipartition theorem. For practical reasons, the thermostat can act always on the underlying explicit identity of the molecules even if they are in the coarse grained region (keeping a double resolution)[15]. The explicit forces are then added up to determine the force acting on the center of mass of the coarse grained molecules. In this way the coarse grained particles have the correct velocity distribution.

Diffusive processes

The application of the AdResS method as reported in the previous sections may lead to the fact that one has different diffusion constants in the atomistic and in the coarse grained region. This will lead to an asymmetric diffusion profile for molecules whose coarse grained representation is much simplified with respect to the atomistic one (for example for water). However, while a faster dynamics of the coarse grained molecules may even be an advantage for sampling purposes, for dynamical analysis this is not ideal. A way to circumvent

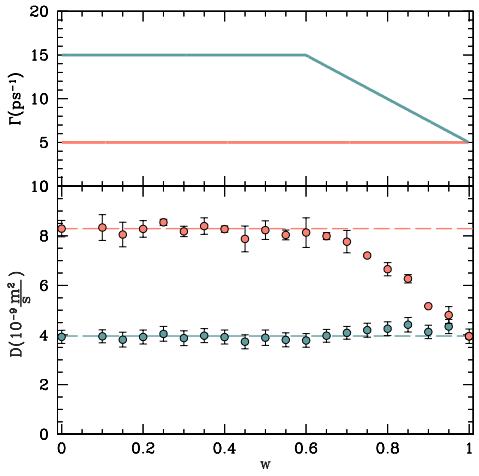


Figure 10. Top figure: The upper curve indicates the dependency of the friction coefficient as a function of the particle identity w when a position-dependent thermostat is used. The straight (lower) curve shows the constant value of the friction coefficient when a regular thermostat is used. Bottom figure: The dots at the upper part indicate the diffusion of the molecules when the regular thermostat is used. The dots of the lower part indicate the diffusion of the molecules when the position-dependent thermostat is used. (Figure was taken from Ref.[29])

this problem is that of slowing down the dynamics of the faster coarse grained molecules. The Langevin thermostat (see above) allows for the changing of the dynamics (and the diffusion constant) by modifying the strength of the friction ζ. As the Langevin thermostat is a local thermostat one can easily make the friction space dependent (or weight dependent). In this case one has to simply tune $\zeta(w)$ in a way that the diffusion constant is the same all over the system. This has been done for the tetrahedral fluid (see Fig. 10).

Recently[41] the DPD thermostat has been extended to change the dynamics of the system; the basic idea is to add an additional friction (and noise) to the transversal degrees of freedom, which allows to conserve hydrodynamics keeping Galilei invariance.

Acknowledgments

We are grateful to Christine Peter and Kurt Kremer for a critical reading of this manuscript and useful suggestions. We would also like to thank Cecilia Clementi, Silvina Matysiak, and Rafael Delgado-Buscalioni for a fruitful collaboration and many discussions on the topics described in this lecture.

References

1. L. Delle Site, C. F. Abrams, A. Alavi, and K. Kremer, *Polymers near Metal Surfaces: Selective Adsorption and Global Conformations*, Phys. Rev. Lett. **89**, 156103, 2002.
2. L. Delle Site, S. Leon, and K. Kremer, *BPA-PC on a Ni(111) Surface: The Interplay between Adsorption Energy and Conformational Entropy for different Chain-End Modifications.*, J. Am. Chem. Soc. **126**, 2944, 2004.

3. O. Alexiadis, V. A. Harmandaris, V. G. Mavrantzas, and L. Delle Site, *Atomistic simulation of alkanethiol self-assembled monolayers on different metal surfaces via a quantum, first-principles parameterization of the sulfur-metal interaction*, J. Phys. Chem. C **111**, 6380, 2007.
4. L. M. Ghiringhelli, B. Hess, N. F. A. van der Vegt, and L. Delle Site, *Competing adsorption between hydrated peptides and water onto metal surfaces: from electronic to conformational properties*, J. Am. Chem. Soc. **130**, 13460, 2008.
5. K. Reuter, D. Frenkel, and M. Scheffler, *The steady state of heterogeneous catalysis, studied by first-principles statistical mechanics*, Phys. Rev. Lett. **93**, 116105, 2004.
6. J. Rogal, K. Reuter, and M. Scheffler, *CO oxidation on Pd(100) at technologically relevant pressure conditions: First-principles kinetic Monte Carlo study*, Phys. Rev. B **77**, 155410, 2008.
7. K. Tarmyshov and F. Müller-Plathe, *Interface between platinum(111) and liquid isopropanol (2-propanol): A model for molecular dynamics studies*, J. Chem. Phys. **126**, 074702, 2007.
8. Gregory A. Voth, editor, in *Coarse Graining of Condensed Phase and Biomolecular Systems*, Chapman and Hall/CRC Press, Taylor and Francis Group, 2008.
9. G. Lu, E. B. Tadmor, and E. Kaxiras, *From electrons to finite elements: A concurrent multiscale approach for metals*, Phys. Rev. B **73**, 024108, 2006.
10. J. Rottler, S. Barsky, and M. O. Robbins, *Cracks and Crazes: On Calculating the Macroscopic Fracture Energy of Glassy Polymers from Molecular Simulations*, Phys. Rev. Lett. **89**, 148304, 2002.
11. G. Csanyi, T. Albaret, M. C. Payne, and A. D. Vita, *"Learn on the Fly": A Hybrid Classical and Quantum-Mechanical Molecular Dynamics Simulation*, Phys. Rev. Lett. **93**, 175503, 2004.
12. A. Laio, J. VandeVondele, and U. Röthlisberger, *A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations*, J. Chem. Phys. **116**, 6941, 2002.
13. D. E. Jiang and E. A. Carter, *First principles assessment of ideal fracture energies of materials with mobile impurities: implications for hydrogen embrittlement of metals*, Acta Materialia **52**, 4801, 2004.
14. G. Lu and E. Kaxiras, *Hydrogen Embrittlement of Aluminum: The Crucial Role of Vacancies*, Phys. Rev. Lett. **94**, 155501, 2005.
15. M. Praprotnik, L. Delle Site, and K. Kremer, *Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly*, J. Chem. Phys. **123**, 224106, 2005.
16. M. Praprotnik, L. Delle Site, and K. Kremer, *Adaptive Resolution Scheme (AdResS) for Efficient Hybrid Atomistic/Mesoscale Molecular Dynamics Simulations of Dense Liquids*, Phys. Rev. E **73**, 066701, 2006 .
17. M. Praprotnik, L. Delle Site, and K. Kremer, *Multiscale Simulation of Soft Matter: From Scale Bridging to Adaptive Resolution*, Annu. Rev. Phys. Chem. **59**, 545, 2008.
18. B. Ensing, S. O. Nielsen, P. B. Moore, M. L. Klein, and M. Parrinello, *Energy conservation in adaptive hybrid atomistic/coarse grain molecular dynamics*, J. Chem. Theor. Comp. **3**, 1100, 2007.
19. A. Heyden, and D. G. Truhlar, *A conservative algorithm for an adaptive change of resolution in mixed atomistic/coarse grained multiscale simulations*, J. Chem. Theor.

Comp. **4**, 217, 2008.
20. A. Alavi, in *Monte Carlo and Molecular Dynamics of Condensed Matter Systems*, chapter 25, page 649. Italian Physical Society, Bologna, 1996.
21. B. Bernu, and D. M. Ceperley, in *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, page 51 NIC series, Volume 10, 2002.
22. S. Nielsen, R. Kapral and G. Ciccotti, *Statistical mechanics of quantum-classical systems*, J. Chem. Phys. **115**, 5805, 2001.
23. M. Praprotnik, K. Kremer, and L. Delle Site, *Adaptive molecular resolution via a continuous change of the phase space dimensionality*, Phys. Rev. E **75**, 017701, 2007.
24. L. Delle Site, *Some fundamental problems for an energy-conserving adaptive-resolution molecular dynamics scheme*, Phys. Rev. E **76**, 047701, 2007.
25. L. Delle Site, *The Adaptive Resolution Simulation method (AdResS): Basic principles and mathematical challenges*, 2008 Reports of the Mathematisches Forschungsinstitut Oberwolfach **21**, 27, 2008.
26. M. Praprotnik, K. Kremer, and L. Delle Site, *Fractional dimensions of phase space variables: A tool for varying the degrees of freedom of a system in a multiscale treatment*, J. Phys. A: Math. Gen. **40**, F281, 2007.
27. M. Praprotnik, L. Delle Site, and K. Kremer, *A macromolecule in a solvent: Adaptive resolution molecular dynamics simulation*, J. Chem. Phys. **126**, 134902, 2007.
28. M. Praprotnik, S. Matysiak, L. Delle Site, K. Kremer and C. Clementi, *Adaptive resolution simulation of liquid water*, J. Phys. Cond. Matt. **19**, 292201, 2007.
29. S. Matysiak, C. Clementi, M. Praprotnik, K. Kremer, and L. Delle Site, *Modeling Diffusive Dynamics in Adaptive Resolution Simulation of Liquid Water*, J. Chem. Phys. **128**, 024503, 2008.
30. R. Delgado-Buscalioni, K. Kremer, and M. Praprotnik, *Concurrent triple-scale simulation of molecular liquids*, J. Chem. Phys. **128**, 114110, 2008.
31. R. Delgado-Buscalioni, K. Kremer, and M. Praprotnik, *Coupling atomistic and continuum hydrodynamics though a mesoscopic model: application to liquid water*, (2008) Submitted.
32. G. De Fabritiis, R. Delgado-Buscalioni, and P. Coveney, *Multiscale Modeling of Liquids with Molecular Specificity*, Phys. Rev. Lett **97**, 134501, 2006.
33. R. Delgado-Buscalioni and G. De Fabritiis, *Embedding Molecular Dynamics within Fluctuating Hydrodynamics in Multiscale Simulations of Liquids*, Phys. Rev. E **76**, 036709, 2007.
34. H.-J. Limbach, A. Arnold, B. A. Mann, and C. Holm, *ESPResSo - An Extensible Simulation Package for Research on Soft Matter Systems*, Comput. Phys. Commun. **174**, 704–727, 2006. `http://www.espresso.mpg.de`
35. D. Reith, M. Puetz, and F. Müller-Plathe, *Deriving effective mesoscale potentials from atomistic simulations* J. Comp. Chem. **24**, 1624–1636, 2003. ,
36. M. Neumann, *Dipole moment fluctuation formulas in computer simulations of polar systems*, Mol. Phys. **50**, 841–858, 1983.
37. P. E. Smith, and W. F. van Gunsteren, *Consistent dielectric properties of the simple point charge and extended simple point charge water models at 277 and 300 K*, J. Chem. Phys. **100**, 3169–3174, 1993.
38. B. Dünweg, *Langevin Methods in B. Dünweg, D. P. Landau, and A. I. Milchev,* Computer simulations of surfaces and interfaces , *proceedings of the NATO Advanced*

Study Institute / Euroconference, Albena, Bulgaria, September 2002, Kluwer Academic Publishers, Dordrecht / Boston / London (2003).

39. *H. Risken,* The Fokker-Planck Equation, *Berlin: Springer Verlag.*
40. *T. Soddemann, B. Dünweg, and K. Kremer,* Dissipative particle dynamics: A useful thermostat for equilibrium and nonequilibrium molecular dynamics simulations, *Phys. Rev. E* **68**, *046702, 2003.*
41. *C. Junghans, M. Praprotnik, and K. Kremer,* Transport properties controlled by a thermostat: An extended dissipative particle dynamics thermostat, *Soft Matter* **4**, *156, 2008.*

Computer Simulations of Systems with Hydrodynamic Interactions: The Coupled Molecular Dynamics – Lattice Boltzmann Approach

Burkhard Dünweg

Max Planck Institute for Polymer Research
Ackermannweg 10, 55128 Mainz, Germany
E-mail: duenweg@mpip-mainz.mpg.de

In soft–matter systems where Brownian constituents are immersed in a solvent, both thermal fluctuations and hydrodynamic interactions are important. The article outlines a general scheme to simulate such systems by coupling Molecular Dynamics for the Brownian particles to a lattice Boltzmann algorithm for the solvent. As an application, the computer simulation of colloidal electrophoresis is briefly discussed.

1 Introduction

Remark: The present contribution intends to just give a very brief overview over the subject matter. The author has recently, together with A. J. C. Ladd, written a 76–page review article[1], to which the interested reader is referred. Detailed explanations and derivations, as well as an extended reference list, can be found there. —

Many soft–matter systems are comprised of Brownian particles immersed in a solvent. Prototypical examples are colloidal dispersions and polymer solutions, where the latter, in contrast to the former, are characterized by non–trivial internal degrees of freedom (here: the many possible conformations of the macromolecule). Fundamental for these systems is the separation of length and time scales between "large and slow" Brownian particles, and "small and fast" solvent particles. "Mesoscopic" simulations focus on the range of length and time scales which are, on the one hand, too small to allow a description just in terms of continuum mechanics of the overall system, but, on the other hand, large enough to allow the replacement of the solvent by a hydrodynamic continuum. This latter approximation is much less severe than one would assume at first glance; detailed Molecular Dynamics simulations have shown that hydrodynamics works as soon as the length scale exceeds a few particle diameters, and the time scale a few collision times.

To simulate such systems consistently, one has to take into account that the length and time scales are so small that thermal fluctuations cannot be neglected. The "Boltzmann number" Bo (a term invented by us) is a useful parameter for quantifying how important fluctuations are. Given a certain spatial resolution b (for example, the lattice spacing of a grid which is used to simulate the fluid dynamics), we may ask ourselves how many solvent particles N_p correspond to the scale b. On average, this is given by $N_p = \rho b^3/m_p$, where ρ is the mass density and m_p the mass of a solvent particle (and we assume a three–

dimensional system). The relative importance of fluctuations is then given by

$$Bo = N_p^{-1/2} = \left(\frac{m_p}{\rho b^3}\right)^{1/2}. \tag{1}$$

It should be noted that for an ideal gas, where the occupation statistics is Poissonian, Bo is just the relative statistical inaccuracy of the random variable N_p. In soft–matter systems, b is usually small enough such that Bo is no longer negligible.

Furthermore, *hydrodynamic interactions* must be modeled. In essence, this term refers to dynamic correlations between the Brownian particles, mediated by fast momentum transport through the solvent. The separation of time scales can be quantified in terms of the so–called Schmidt number

$$Sc = \frac{\eta_{kin}}{D}, \tag{2}$$

where $\eta_{kin} = \eta/\rho$ is the kinematic viscosity (ratio of dynamic shear viscosity η and mass density ρ) of the fluid, measuring how quickly momentum propagates diffusively through the solvent, and D is the diffusion constant of the particles. Typically, in a dense fluid $Sc \sim 10^2 \ldots 10^3$ for the solvent particles, while for large Brownian particles Sc is even much larger. Finally, we may also often assume that the solvent dynamics is in the creeping–flow regime, i. e. that the Reynolds number

$$Re = \frac{ul}{\eta_{kin}}, \tag{3}$$

where u denotes the velocity of the flow and l its typical size, is small. This is certainly true as long as the system is not driven strongly out of thermal equilibrium.

These considerations lead to the natural (but, in our opinion, not always correct) conclusion that the method of choice to simulate such systems is Brownian Dynamics[2]. Here the Brownian particles are displaced under the influence of particle–particle forces, hydrodynamic drag forces (calculated from the particle positions), and stochastic forces representing the thermal noise. However, the technical problems to do this efficiently for a large number N of Brownian particles are substantial. The calculation of the drag forces involves the evaluation of the hydrodynamic Green's function, which depends on the boundary conditions, and has an intrinsically long–range nature (such that all particles interact with each other). Furthermore, these drag terms also determine the correlations in the stochastic displacements, such that the generation of the stochastic terms involves the calculation of the matrix square root of a $3N \times 3N$ matrix. Recently, there has been substantial progress in the development of fast algorithms[3]; however, currently there are only few groups who master these advanced and complicated techniques. Apart from this, the applicability is somewhat limited, since the Green's function must be re–calculated for each new boundary condition, and its validity is questionable if the system is put under strong nonequilibrium conditions like, e. g., a turbulent flow — it should be noted that the Green's function is calculated for low–Re hydrodynamics.

Therefore, many soft–matter researchers have rather chosen the alternative approach, which is to simulate the system including the solvent degrees of freedom, with explicit momentum transport. The advantage of this is a simple algorithm, which scales linearly with the number of Brownian particles, and is easily parallelizable, due to its locality. The disadvantage, however, is that one needs to simulate many more degrees of freedom than

those in which one is genuinely interested — *and* to do this on the short inertial time scales in which one is not interested either. It is clear that such an approach involves essentially Molecular Dynamics (MD) for the Brownian particles.

Many ways are possible how to simulate the solvent degrees of freedom, and how to couple them to the MD part. It is just the universality of hydrodynamics that allows us to invent many models which all will produce the correct physics. The requirements are rather weak — the solvent model has to just be compatible with Navier–Stokes hydrodynamics on the macroscopic scale. Particle methods include Dissipative Particle Dynamics (DPD) and Multi–Particle Collision Dynamics (MPCD)[4], while lattice methods involve the direct solution of the Navier–Stokes equation on a lattice, or lattice Boltzmann (LB). The latter is a method with which we have made quite good experience, both in terms of efficiency and versatility. The efficiency comes from the inherent ease of memory management for a lattice model, combined with ease of parallelization, which comes from the high degree of locality: Essentially an LB algorithm just shifts populations on a lattice, combined with collisions, which however only happen locally on a single lattice site. The coupling to the Brownian particles (simulated via MD) can either be done via boundary conditions, or via an interpolation function that introduces a *dissipative* coupling between particles and fluid. In this article, we will focus on the latter method.

2 Coupling Scheme

As long as we view LB as just a solver for the Navier–Stokes equation, we may write down the equations of motion for the coupled system as follows:

$$\frac{d}{dt}\vec{r}_i = \frac{1}{m_i}\vec{p}_i, \tag{4}$$

$$\frac{d}{dt}\vec{p}_i = \vec{F}_i^c + \vec{F}_i^d + \vec{F}_i^f, \tag{5}$$

$$\partial_t \rho + \partial_\alpha j_\alpha = 0, \tag{6}$$

$$\partial_t j_\alpha + \partial_\beta \pi^E_{\alpha\beta} = \partial_\beta \eta_{\alpha\beta\gamma\delta}\partial_\gamma u_\delta + f^h_\alpha + \partial_\beta \sigma^f_{\alpha\beta}. \tag{7}$$

Here, $\vec{r}_i$, $\vec{p}_i$ and m_i are the positions, momenta, and masses of the Brownian particles, respectively. The forces $\vec{F}_i$ acting on the particles are conservative (c, i. e. coming from the interparticle potential), dissipative (d), and fluctuating (f). The equations of motion for the fluid have been written in tensor notation, where Greek indexes denote Cartesian components, and the Einstein summation convention is used. The first equation describes mass conservation; the mass flux $\rho\vec{u}$, where $\vec{u}$ is the flow velocity, is identical to the momentum density $\vec{j}$. The last equation describes the time evolution of the fluid momentum density. In the absence of particles, the fluid momentum is conserved. This part is described via the stress tensor, which in turn is decomposed into the conservative Euler stress $\pi^E_{\alpha\beta}$, the dissipative stress $\eta_{\alpha\beta\gamma\delta}\partial_\gamma u_\delta$, and the fluctuating stress $\sigma^f_{\alpha\beta}$. The influence of the particles is described via an external force density $\vec{f}^h$.

The coupling to a particle i is introduced via an interpolation procedure where first the flow velocities from the surrounding sites are averaged over to yield the flow velocity right

at the position of i. In the continuum limit, this is written as

$$\vec{u}_i \equiv \vec{u}(\vec{r}_i) = \int d^3\vec{r}\,\Delta(\vec{r},\vec{r}_i)\vec{u}(\vec{r}), \tag{8}$$

where $\Delta(\vec{r},\vec{r}_i)$ is a weight function with compact support, satisfying

$$\int d^3\vec{r}\Delta(\vec{r},\vec{r}_i) = 1. \tag{9}$$

Secondly, each particle is assigned a phenomenological friction coefficient Γ_i, and this allows us to calculate the friction force on particle i:

$$\vec{F}_i^d = -\Gamma_i\left(\frac{\vec{p}_i}{m_i} - \vec{u}_i\right). \tag{10}$$

A Langevin noise term $\vec{F}_i^f$ is added to the particle equation of motion, in order to compensate the dissipative losses that come from $\vec{F}_i^d$. $\vec{F}_i^f$ satisfies the standard fluctuation–dissipation relation

$$\left\langle F_{i\alpha}^f \right\rangle = 0, \tag{11}$$

$$\left\langle F_{i\alpha}^f(t)\, F_{j\beta}^f(t')\right\rangle = 2k_BT\Gamma_i\delta_{ij}\delta_{\alpha\beta}\delta(t-t'), \tag{12}$$

where T is is the absolute temperature and k_B the Boltzmann constant. While the conservative forces $\vec{F}_i^c$ conserve the total momentum of the particle system, as a result of Newton's third law, the dissipative and fluctuating terms ($\vec{F}_i^d$ and $\vec{F}_i^f$) do not. The associated momentum transfer must therefore have come from the fluid. The overall momentum must be conserved, however. This means that the force term entering the Navier–Stokes equation must just balance these forces. One easily sees that the choice

$$\vec{f}^h(\vec{r}) = -\sum_i\left(\vec{F}_i^d + \vec{F}_i^f\right)\Delta(\vec{r},\vec{r}_i) \tag{13}$$

satisfies this criterion. It should be noted that we use the *same* weight function to interpolate the forces back onto the fluid; this is necessary to satisfy the fluctuation–dissipation theorem for the overall system, i. e. to simulate a well–defined constant–temperature ensemble. The detailed proof of the thermodynamic consistency of the procedure can be found in Ref. 1.

We still need to specify the remaining terms in the Navier–Stokes equation. The viscosity tensor $\eta_{\alpha\beta\gamma\delta}$ describes an isotropic Newtonian fluid:

$$\eta_{\alpha\beta\gamma\delta} = \eta\left(\delta_{\alpha\gamma}\delta_{\beta\delta} + \delta_{\alpha\delta}\delta_{\beta\gamma} - \frac{2}{3}\delta_{\alpha\beta}\delta_{\gamma\delta}\right) + \eta_v\delta_{\alpha\beta}\delta_{\gamma\delta}, \tag{14}$$

with shear and bulk viscosities η and η_v. This tensor also appears in the covariance matrix of the fluctuating (Langevin) stress $\sigma_{\alpha\beta}^f$:

$$\left\langle \sigma_{\alpha\beta}^f \right\rangle = 0, \tag{15}$$

$$\left\langle \sigma_{\alpha\beta}^f(\vec{r},t)\,\sigma_{\gamma\delta}^f(\vec{r}',t')\right\rangle = 2k_BT\eta_{\alpha\beta\gamma\delta}\delta(\vec{r}-\vec{r}')\,\delta(t-t'). \tag{16}$$

Finally, the Euler stress

$$\pi^E_{\alpha\beta} = p\delta_{\alpha\beta} + \rho u_\alpha u_\beta \tag{17}$$

describes the equation of state of the fluid (p is the thermodynamic pressure), and convective momentum transport.

3 Low Mach Number Physics

At this point an important simplification can be made. The equation of state only matters for flow velocities u that are comparable with the speed of sound c_s, i. e. for which the Mach number

$$Ma = \frac{u}{c_s} \tag{18}$$

is large. In the low Mach number regime, the flow may be considered as effectively incompressible (although no incompressibility constraint is imposed in the algorithm). The Mach number should not be confused with the Reynolds number Re, which rather measures whether inertial effects are important. Now it turns out that essentially all soft–matter applications "live" in the low–Ma regime. Furthermore, large Ma is anyways inaccessible to the LB algorithm, since it provides only a finite set of lattice velocities — and these essentially determine the value of c_s. In other words, the LB algorithm simply cannot realistically represent flows whose velocity is not small compared to c_s. For this reason, the details of the equation of state do not matter, and therefore one chooses the system that is by far the easiest — the ideal gas. Here the equation of state for a system at temperature T may be written as

$$k_B T = m_p c_s^2. \tag{19}$$

In the D3Q19 model (the most popular standard LB model in three dimensions, using nineteen lattice velocities, see below) it turns out that the speed of sound is given by

$$c_s^2 = \frac{1}{3}\frac{b^2}{h^2}, \tag{20}$$

where b is the lattice spacing and h the time step. Therefore the Boltzmann number can also be written as

$$Bo = \left(\frac{m_p}{\rho b^3}\right)^{1/2} = \left(\frac{3k_B T h^2}{\rho b^5}\right)^{1/2}. \tag{21}$$

4 Lattice Boltzmann 1: Statistical Mechanics

The lattice Boltzmann algorithm starts from a regular grid with sites $\vec{r}$ and lattice spacing b, plus a time step h. We then introduce a small set of velocities $\vec{c}_i$ such that $\vec{c}_i h$ connects two nearby lattice sites on the grid. In the D3Q19 model, the lattice is simple cubic, and the nineteen velocities correspond to the six nearest and twelve next–nearest neighbors, plus a zero velocity. On each lattice site $\vec{r}$ at time t, there are nineteen populations $n_i(\vec{r}, t)$.

Each population is interpreted as the mass density corresponding to velocity $\vec{c}_i$. The total mass and momentum density are therefore given by

$$\rho(\vec{r},t) = \sum_i n_i(\vec{r},t), \tag{22}$$

$$\vec{j}(\vec{r},t) = \sum_i n_i(\vec{r},t)\vec{c}_i, \tag{23}$$

such that the flow velocity is obtained via $\vec{u} = \vec{j}/\rho$. The number of "lattice Boltzmann particles" which correspond to n_i is given by

$$\nu_i = \frac{n_i b^3}{m_p} \equiv \frac{n_i}{\mu}, \tag{24}$$

where m_p is the mass of a lattice Boltzmann particle, and μ the corresponding mass density. It should be noted that μ is a measure of the thermal fluctuations in the system, since, according to Eq. 21, one has $Bo^2 = \mu/\rho$.

If we now assume a "velocity bin" i to be in thermal contact with a large reservoir of particles, the probability density for ν_i is Poissonian. Furthermore, if we assume that the "velocity bins" are statistically independent, but take into account that mass and momentum density are fixed (these variables are conserved quantities during an LB collision step and should therefore be handled like conserved quantities in a microcanonical ensemble), we find

$$P\left(\{\nu_i\}\right) \propto \left(\prod_i \frac{\bar{\nu}_i^{\nu_i}}{\nu_i!} e^{-\bar{\nu}_i}\right) \delta\left(\mu \sum_i \nu_i - \rho\right) \delta\left(\mu \sum_i \nu_i \vec{c}_i - \vec{j}\right). \tag{25}$$

for the probability density of the variables ν_i. This must be viewed as the statistics which describes the local (single–site) equilibrium under the condition of fixed values of the hydrodynamic variables ρ and $\vec{j}$. The parameter $\bar{\nu}_i$ is the mean occupation imposed by the reservoir, and we assume that it is given by

$$\bar{\nu}_i = a^{c_i} \frac{\rho}{\mu}, \tag{26}$$

where $a^{c_i} > 0$ is a weight factor corresponding to the neighbor shell with speed c_i.

From normalization and cubic symmetry we know that the low–order velocity moments of the weights must have the form

$$\sum_i a^{c_i} = 1, \tag{27}$$

$$\sum_i a^{c_i} c_{i\alpha} = 0, \tag{28}$$

$$\sum_i a^{c_i} c_{i\alpha} c_{i\beta} = \sigma_2\, \delta_{\alpha\beta}, \tag{29}$$

$$\sum_i a^{c_i} c_{i\alpha} c_{i\beta} c_{i\gamma} = 0, \tag{30}$$

$$\sum_i a^{c_i} c_{i\alpha} c_{i\beta} c_{i\gamma} c_{i\delta} = \kappa_4\, \delta_{\alpha\beta\gamma\delta} + \sigma_4 \left(\delta_{\alpha\beta}\delta_{\gamma\delta} + \delta_{\alpha\gamma}\delta_{\beta\delta} + \delta_{\alpha\delta}\delta_{\beta\gamma}\right), \tag{31}$$

where σ_2, σ_4, κ_4 are yet undetermined constants, while $\delta_{\alpha\beta\gamma\delta}$ is unity if all four indexes are the same and zero otherwise.

Employing Stirling's formula for the factorial, it is straightforward to find the set of populations n_i^{eq} which maximizes P under the constraints of given ρ and $\vec{j}$. Up to second order in u (low Mach number!) the solution is given by

$$n_i^{eq} = \rho a^{c_i} \left(1 + \frac{\vec{u} \cdot \vec{c}_i}{\sigma_2} + \frac{(\vec{u} \cdot \vec{c}_i)^2}{2\sigma_2^2} - \frac{u^2}{2\sigma_2} \right). \tag{32}$$

The low–order moments of the equilibrium populations are then given by

$$\sum_i n_i^{eq} = \rho, \tag{33}$$

$$\sum_i n_i^{eq} c_{i\alpha} = j_\alpha, \tag{34}$$

$$\sum_i n_i^{eq} c_{i\alpha} c_{i\beta} = \rho c_s^2 \delta_{\alpha\beta} + \rho u_\alpha u_\beta. \tag{35}$$

The first two equations are just the imposed constraints, while the last one (meaning that the second moment is just the hydrodynamic Euler stress) follows from imposing two additional conditions, which is to choose the weights a^{c_i} such that they satisfy $\kappa_4 = 0$ and $\sigma_4 = \sigma_2^2$. From the Chapman–Enskog analysis of the LB dynamics (see below) it follows that the asymptotic behavior in the limit of large length and time scales is compatible with the Navier–Stokes equation only if Eq. 35 holds, and this in turn is only possible if the abovementioned isotropy conditions are satisfied. Together with the normalization condition, we thus obtain a set of three equations for the a^{c_i}. Therefore at least three neighbor shells are needed to satisfy these conditions, and this is the reason for choosing a nineteen–velocity model. For D3Q19, one thus obtains $a^{c_i} = 1/3$ for the zero velocity, $1/18$ for the nearest neighbors, and $1/36$ for the next–nearest neighbors. Furthermore, one finds $c_s^2 = \sigma_2 = (1/3)b^2/h^2$.

For the fluctuations around the most probable populations n_i^{eq},

$$n_i^{neq} = n_i - n_i^{eq}, \tag{36}$$

we employ a saddle–point approximation and approximate u by zero. This yields

$$P\left(\{n_i^{neq}\}\right) \propto \exp\left(- \sum_i \frac{(n_i^{neq})^2}{2\mu\rho a^{c_i}} \right) \delta\left(\sum_i n_i^{neq} \right) \delta\left(\sum_i \vec{c}_i \, n_i^{neq} \right). \tag{37}$$

We now introduce normalized fluctuations via

$$\hat{n}_i^{neq} = \frac{n_i^{neq}}{\sqrt{\mu\rho a^{c_i}}} \tag{38}$$

and transform to normalized "modes" (symmetry–adapted linear combinations of the n_i, see Ref. 1) $\hat{m}_k^{neq}$ via an orthonormal transformation $\hat{e}_{ki}$:

$$\hat{m}_k^{neq} = \sum_i \hat{e}_{ki} \hat{n}_i^{neq}, \tag{39}$$

$k = 0, \ldots, 18$, and obtain

$$P\left(\{m_k\}\right) \propto \exp\left(-\frac{1}{2}\sum_{k\geq 4} m_k^2\right). \tag{40}$$

It should be noted that the modes number zero to three have been excluded; they are just the conserved mass and momentum densities.

5 Lattice Boltzmann 2: Stochastic Collisions

A collision step consists of re-arranging the set of n_i on a given lattice site such that both mass and momentum are conserved. Since the algorithm should simulate thermal fluctuations, this should be done in a way that is (i) stochastic and (ii) consistent with the developed statistical–mechanical model. This is straightforwardly imposed by requiring that the collision is nothing but a Monte Carlo procedure, where a Monte Carlo step transforms the pre–collisional set of populations, n_i, to the post–collisional one, $n_i^\star$. Consistency with statistical mechanics can be achieved by requiring that the Monte Carlo update satisfies the condition of detailed balance. Most easily this is done in terms of the normalized modes $\hat{m}_k$, which we update according to the rule ($k \geq 4$)

$$\hat{m}_k^\star = \gamma_k \hat{m}_k + \sqrt{1-\gamma_k^2}\, r_k. \tag{41}$$

Here the γ_k are relaxation parameters with $-1 < \gamma_k < 1$, and the r_k are statistically independent Gaussian random numbers with zero mean and unit variance. Mass and momentum are automatically conserved since the corresponding modes are not updated. Comparison with Eq. 40 shows that the procedure indeed does satisfy detailed balance. The parameters γ_k can in principle be chosen at will; however, they should be compatible with symmetry. For example, mode number four corresponds to the bulk stress, with a relaxation parameter γ_b, while modes number five to nine correspond to the five shear stresses, which form a symmetry multiplet. Therefore one must choose $\gamma_5 = \ldots = \gamma_9 = \gamma_s$. For the remaining kinetic modes one often uses $\gamma_k = 0$ for simplicity, but this is not necessary.

6 Lattice Boltzmann 3: Chapman–Enskog Expansion

The actual LB algorithm now consists of alternating collision and streaming steps, as summarized in the LB equation (LBE):

$$n_i(\vec{r} + \vec{c}_i h, t + h) = n_i^\star(\vec{r}, t) = n_i(\vec{r}, t) + \Delta_i\left\{n_i(\vec{r}, t)\right\}. \tag{42}$$

The populations are first re–arranged on the lattice site; this is described by the so–called "collision operator" Δ_i. The resulting post–collisional populations $n_i^\star$ are then propagated to the neighboring sites, as expressed by the left hand side of the equation. After that, the next collision step is done, etc.. The collision step may include momentum transfer as a result of external forces (for details, see Ref. 1); apart from that, it is just given by the update procedure outlined in the previous section.

A convenient way to find the dynamic behavior of the algorithm on large length and time scales is a multi–time–scale analysis. One introduces a "coarse grained ruler" by transforming from the original coordinates $\vec{r}$ to new coordinates $\vec{r}_1$ via

$$\vec{r}_1 = \epsilon \vec{r}, \tag{43}$$

where ϵ is a dimensionless parameter with $0 < \epsilon \ll 1$. The rationale behind this is the fact that any "reasonable" value for the scale r_1 will automatically force r to be large. In other words: By considering the limit $\epsilon \to 0$ we automatically focus our attention on large length scales. The same is done for the time; however, here we introduce *two* scales via

$$t_1 = \epsilon t \tag{44}$$

and

$$t_2 = \epsilon^2 t. \tag{45}$$

The reason for this is that one needs to consider both wave–like phenomena, which happen on the t_1 time scale (i. e. the real time is moderately large), and diffusive processes (where the real time is *very* large). We now write the LB variables as a function of $\vec{r}_1, t_1, t_2$ instead of $\vec{r}, t$. Since changing ϵ at fixed $\vec{r}_1$ changes $\vec{r}$ and thus n_i, we must take into account that the LB variables depend on ϵ:

$$n_i = n_i^{(0)} + \epsilon n_i^{(1)} + \epsilon^2 n_i^{(2)} + O(\epsilon^3). \tag{46}$$

The same is true for the collision operator:

$$\Delta_i = \Delta_i^{(0)} + \epsilon \Delta_i^{(1)} + \epsilon^2 \Delta_i^{(2)} + O(\epsilon^3). \tag{47}$$

In terms of the new variables, the LBE is written as

$$n_i(\vec{r}_1 + \epsilon \vec{c}_i h, t_1 + \epsilon h, t_2 + \epsilon^2 h) - n_i(\vec{r}_1, t_1, t_2) = \Delta_i. \tag{48}$$

Now, one systematically Taylor–expands the equation up to order ϵ^2. Sorting by order yields a hierarchy of LBEs of which one takes the zeroth, first, and second velocity moment. Systematic analysis of this set of moment equations (for details, see Ref. 1) shows that the LB procedure, as it has been developed in the previous sections, indeed yields the fluctuating Navier–Stokes equations in the asymptotic $\epsilon \to 0$ limit — however only for low Mach numbers; in the high Mach number regime, where terms of order u^3/c_s^3 can no longer be neglected, the dynamics definitely deviates from Navier–Stokes.

In particular, this analysis shows that the zeroth–order populations must be identified with n_i^{eq}, and that it is *necessary* that this "encodes" the Euler stress via suitably chosen weights a^{c_i}. Furthermore, one finds explicit expressions for the viscosities:

$$\eta = \frac{h\rho c_s^2}{2} \frac{1+\gamma_s}{1-\gamma_s}, \tag{49}$$

$$\eta_b = \frac{h\rho c_s^2}{3} \frac{1+\gamma_b}{1-\gamma_b}. \tag{50}$$

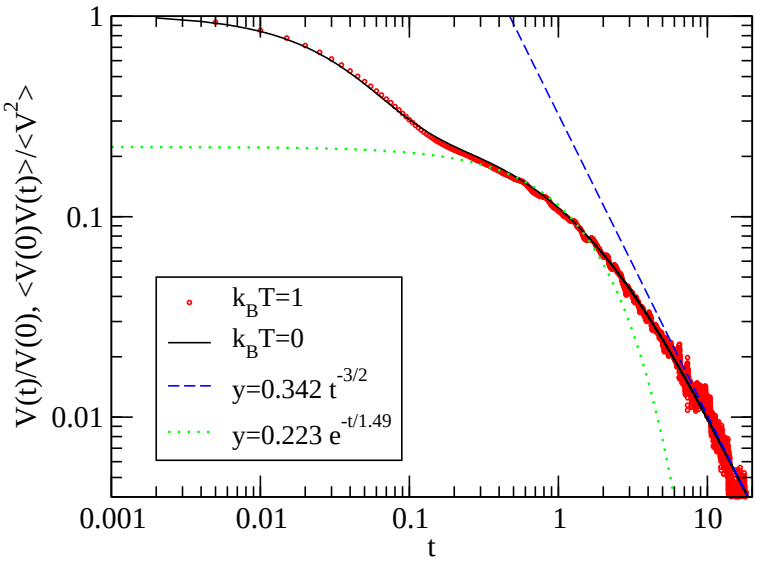


Figure 1. (From Ref. 5) Velocity autocorrelation function of a single colloidal sphere, normalized by the initial value, in thermal equilibrium. The velocity is here defined as the center–of–mass velocity of the particles which form the sphere. $\langle v^2 \rangle$, i. e. the $t = 0$ value of the unnormalized function, is therefore given by the equipartition theorem of statistical mechanics. For larger times, the surface particles become more and more coupled to the fluid inside the sphere, and thus the effective mass of the sphere increases. This is the reason for the first initial decay before a plateau is reached. After that, the function decays according to the famous $t^{-3/2}$ long–time tail. Finally, the particle becomes coupled to the whole fluid in the whole simulation box and the behavior becomes dominated by this finite–size effect. For comparison, the figure also shows the decay of the colloid velocity in a *deterministic* computer experiment, where the noise amplitude for both the particle dynamics and the LB degrees of freedom has been set to zero, and the particle was "kicked" at $t = 0$. This function has been normalized by the initial value, too. According to linear response theory, both curves must coincide, which they do.

7 Example: Dynamics of Charged Colloids

The coupling scheme that has been described in this article is particularly useful for immersed particles with internal degrees of freedom, like flexible polymer chains, or membranes. It can also be applied to systems whose immersed particles are "hard" (for example, colloidal spheres), although the alternative approach by Ladd (see Ref. 1) that models the particles as rigid bodies interacting with the LB fluid via boundary conditions is probably slightly more efficient. Nevertheless, for reasons of easy program development it makes sense to use the same scheme for both flexible and rigid systems. In what follows, some results for a colloidal system shall be presented, in order to demonstrate that and how the method works.

In Ref. 5 we have developed the so–called "raspberry model" for a colloidal sphere. Since the model is intended for charged systems with explicit (salt and counter) ions, it should take into account (at least to some degree) the size difference between colloids and ions. Therefore the colloid is, in terms of linear dimension, roughly 6–7 times larger than the small particles. The LB lattice spacing is chosen as identical to the small ion diameter. This is combined with a linear force interpolation to the nearest neighbor sites.

A larger lattice spacing would result in a rather coarse description of the hydrodynamic interactions, while a yet smaller spacing would result in a large computational overhead. In this context, it should be noted that one would obtain an ill–defined model with infinite particle mobility if one would let the lattice spacing tend to zero, while sticking to the nearest–neighbor interpolation scheme[1]. This is due to the fact that the effective long–time mobility that results from the dissipative coupling is not given by $1/\Gamma$, but rather by

$$\frac{1}{\Gamma_{eff}} = \frac{1}{\Gamma} + \frac{1}{g\eta\sigma}, \tag{51}$$

where σ is the range of the interpolation function and g a numerical prefactor. Therefore, one needs to keep the range of the interpolation function constant, which would involve more and more effort if one would insist on $b \to 0$. Within limits, it is of course possible to compensate the effects of a change of σ by suitably re–adjusting Γ — only the long–time value Γ_{eff} is of real physical significance.

In principle, it would therefore be possible to model a colloidal sphere by a particle which exhibits a suitably chosen excluded–volume interaction for the other (small or large) particles, plus a suitably adjusted large value of the interpolation range σ, which essentially plays the role of a Stokes radius. However, such a model would not describe the rotational degrees of freedom, and these are important. For this reason, we rather model the colloid as a large sphere, around which we wrap a two–dimensional network of small particles (same size as the ions) which are connected via springs. Only the surface particles are coupled dissipatively to the LB fluid. Figures 1 and 2 show that the model behaves exactly as one would expect from hydrodynamics and linear response theory. Figure 1 shows the particle velocity autocorrelation function, from which one obtains, via integration, the translational (or self) diffusion coefficient D^S:

$$D^S = \frac{1}{3}\int_0^\infty dt\ \langle \vec{v}(t)\cdot\vec{v}(0)\rangle\,. \tag{52}$$

In an infinite hydrodynamic continuum, Stokes' law results in the prediction $D^S = k_BT/(6\pi\eta R)$ for a sphere of radius R. Indeed, this is what one finds in that limit. However, for (cubic) simulation boxes of finite linear dimension L, the diffusion constant is systematically smaller, as a result of the hydrodynamic interactions with the periodic images:

$$D^S = \frac{k_BT}{6\pi\eta R} - 2.837\frac{k_BT}{6\pi\eta L}. \tag{53}$$

This is an analytic result, where higher–order terms in the L^{-1} expansion have been neglected. Figure 2 shows that this prediction is nicely reproduced. Furthermore, the rotational diffusion constant, which can be obtained by integrating the angular–velocity autocorrelation function,

$$D^R = \frac{1}{3}\int_0^\infty dt\ \langle \vec{\omega}(t)\cdot\vec{\omega}(0)\rangle\,, \tag{54}$$

exhibits a similar $1/L$ finite size effect; the asymptotic value $k_BT/(8\pi\eta R^3)$ is only reached for infinite system size. As Figure 2 shows, this prediction is reproduced as well.

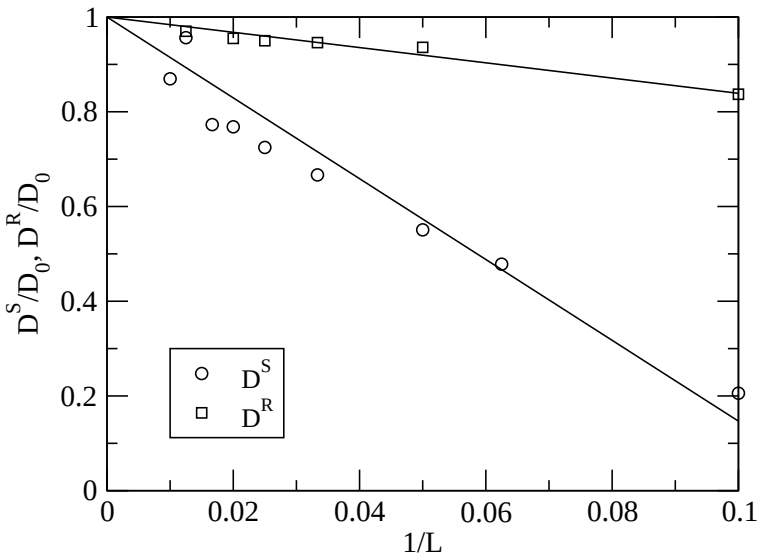


Figure 2. (From Ref. 5) Translational (D^S) and rotational (D^R) diffusion coefficient, normalized by the asymptotic infinite–system value, as a function of inverse system size $1/L$. The straight line is a fit for D^R, while it is the analytical prediction (see text) for D^S.

Electrokinetic phenomena can be investigated by supplying a charge to the central colloidal sphere, and by adding ions such that the total system is charge–neutral. We have studied the electrophoretic mobility, i. e. the response to an external electric field E:

$$\mu = \frac{v}{eE}, \tag{55}$$

where v is the colloid drift velocity and e the elementary charge. The simplest case is to simulate just a single colloid with charge Ze in a cubic box, and to add Z monovalent counterions to compensate the colloidal charge (i. e. no further salt ions are added). This corresponds to a system with a finite volume fraction (one colloid per box). It should be noted that one should *not* consider the limit where this system is being put into larger and larger boxes: In that case, the ions would simply "evaporate", and one would obtain a trivial value for μ that is just given by Stokes' law.

Usually the mobility is given in dimensionless units: The so–called reduced mobility μ_{red} is obtained by normalizing with a Stokes mobility, using the Bjerrum length l_B as the underlying length scale:

$$\mu_{red} = 6\pi\eta l_B \mu, \tag{56}$$

$$l_B = \frac{e^2}{4\pi\varepsilon k_B T}, \tag{57}$$

where ε is the fluid's dielectric constant.

Fortunately, μ is subject to a much smaller finite size effect than the diffusion constant. This has been checked by simulations, see Fig. 3. The reason for this behavior is the fact that the electric field does not exert a net force on the overall system, due to charge

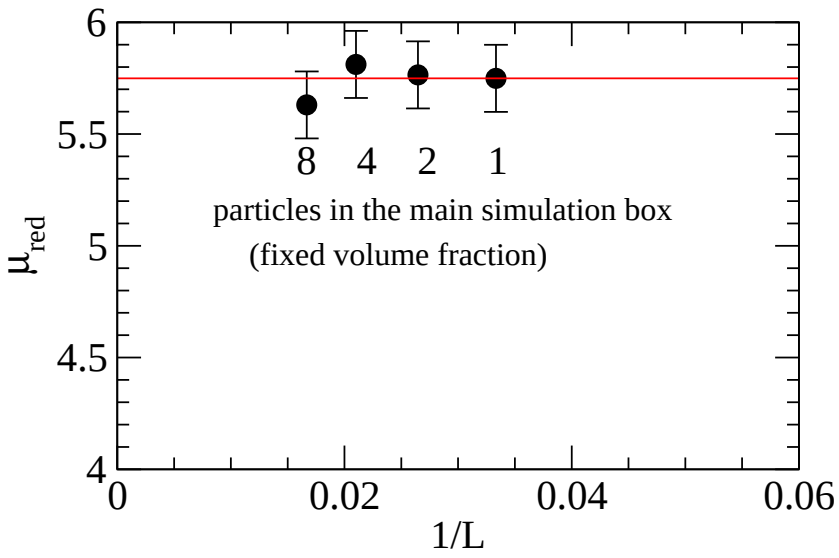


Figure 3. Reduced electrophoretic mobility as a function of inverse system size $1/L$. In order to keep conditions constant (i. e. constant colloidal volume fraction, and constant ion concentration), the box size was systematically increased, while at the same time more and more colloids (up to eight), together with their compensating ions, were put into the box. Within our resolution, no finite size effect could be detected.

neutrality. In other words: The field induces two electric currents in opposite direction. These currents, in turn, induce hydrodynamic flows. These flows, however, cancel each other exactly in leading order. Therefore the hydrodynamic interactions with the periodic images are weak. This should be contrasted with the diffusion constant, which corresponds to the response to an external gravitational field. The latter *does* exert a net force on the overall system, and hence one obtains a large–scale flow decaying like the inverse distance from the colloid. This $1/r$ flow field is exactly the reason for the $1/L$ finite–size effect in the diffusion constant as shown in Fig. 2.

The electrophoretic mobility may be obtained by either applying an electric field, and measuring the drift velocity, or by Green–Kubo integration[6], where a system in strict thermal equilibrium is studied:

$$k_B T \mu = \frac{1}{3} \sum_i z_i \int_0^\infty dt \ \langle \vec{v}_i(0) \cdot \vec{v}_0(t) \rangle \,, \tag{58}$$

where the index i denotes particle number i, and z_i is its valence. Particle number zero is the colloid whose response to the electric field is considered. The nonequilibrium approach is hampered by the fact that, for reasonable electric field values, the response is quite typically in the nonlinear regime (mainly as a result of charge–cloud stripping). Therefore, one needs to extrapolate to zero driving. In contrast, the Green–Kubo value is, *per definition*, the linear–response result. Figure 4 shows that the two approaches yield the same result.

Further results that have been obtained with this model include a study of the concentration dependence of μ, both in terms of colloid volume fraction of a salt–free system, and in terms of salt concentration at fixed colloid concentration. Without going into further de-

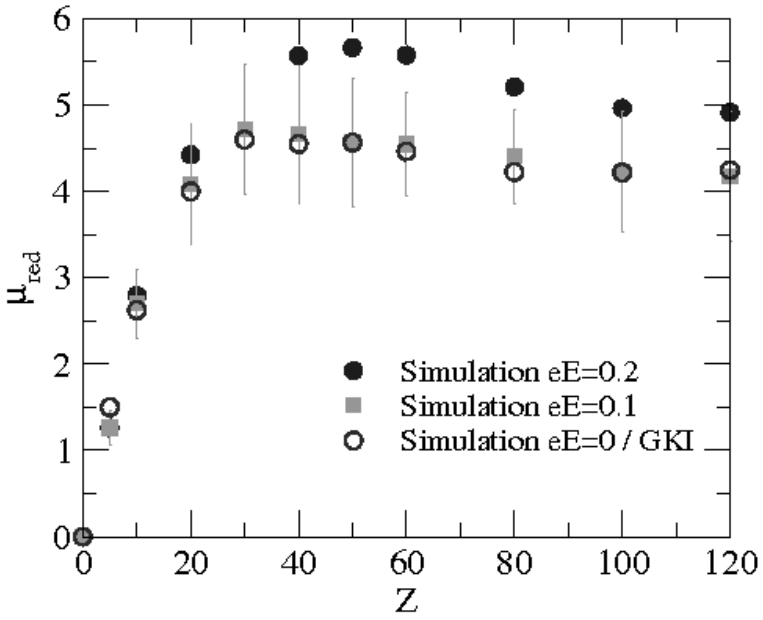


Figure 4. (From Ref. 6) Reduced electrophoretic mobility for the single–colloid system described in the text, as a function of the colloid's charge Z, comparing nonequilibrium with Green–Kubo integration (GKI) results. The mobility first increases, since the force is proportional to the charge. However, for larger Z values it saturates, indicating that more and more ions condense at the colloid's surface, such that the effective charge does not change. For $eE = 0.2$, nonlinear effects lead to an increased mobility, while $eE = 0.1$ is still in the linear–response regime, as demonstrated by the comparison with the equilibrium data.

tails, it should just be mentioned that the reduced–mobility data can be nicely rationalized in terms of a scaling theory[6] which then allows a favorable comparison with experimental results[7].

Of course, this is not the only example where the coupled MD–LB approach has helped to understand the dynamics of soft matter. Other examples include the dynamics of polymers and neutral colloids in both equilibrium and nonequilibrium situations; these have been outlined in Ref. 1. Further simulations will follow in the future, and it seems that the method is gaining popularity in the soft–matter community.

References

1. B. Dünweg and A. J. C. Ladd, "Lattice Boltzmann simulations of soft matter systems", Advances in Polymer Science **221**, 89 2009.
2. G. Nägele, "Brownian dynamics simulations", in: Computational Condensed Matter Physics, S. Blügel, G. Gompper, E. Koch, H. Müller-Krumbhaar, R. Spatschek, and R. G. Winkler, (Eds.). Forschungszentrum Jülich, Jülich, 2006.
3. A. J. Banchio and J. F. Brady, *Accelerated Stokesian dynamics: Brownian motion*, Journal of Chemical Physics, **118**, 10323, 2003.

4. M. Ripoll, “Mesoscale hydrodynamics simulations”, in: Computational Condensed Matter Physics, S. Blügel, G. Gompper, E. Koch, H. Müller-Krumbhaar, R. Spatschek, and R. G. Winkler, (Eds.). Forschungszentrum Jülich, Jülich, 2006.
5. V. Lobaskin and B. Dünweg, *A new model for simulating colloidal dynamics*, New Journal of Physics, **6**, 54, 2004.
6. B. Dünweg, V. Lobaskin, K. Seethalakshmy-Hariharan, and C. Holm, *Colloidal electrophoresis: Scaling analysis, Green-Kubo relation, and numerical results*, Journal of Physics: Condensed Matter, **20**, 404214, 2008.
7. V. Lobaskin, B. Dünweg, M. Medebach, T. Palberg, and C. Holm, *Electrophoresis of colloidal dispersions in the low–salt regime*, Physical Review Letters, **98**, 176105, 2007.

De Novo Protein Folding with Distributed Computational Resources

Timo Strunk[1], Abhinav Verma[2], Srinivasa Murthy Gopal[3], Alexander Schug[4], Konstantin Klenin[1], and Wolfgang Wenzel[1,5]

[1] Forschungszentrum Karlsruhe
Institute for Nanotechnology , P.O. Box 3640, 76021 Karlsruhe, Germany
E-mail: wenzel@int.fzk.de

[2] Centro de Investigaciones Biologicas
Ramiro de Maeztu 9, 28040 Madrid, Spain

[3] Michigan State University
Department of Biochemistry & Molecular Biology
East Lansing, MI 48824-1319, USA

[4] Center for Theoretical Biological Physics (CTBP)
UCSD, La Jolla, CA 92093, USA

[5] DFG Center for Functional Nanotechnology
Karlsruhe Institute for Technology
76131 Karlsruhe, Germany

Proteins constitute a major part of the machinery of all cellular life. While sequence information of many proteins is readily available, the determination of protein three-dimensional structure is much more involved. Computational methods increasingly contribute to elucidate protein structure, conformational change and biological function. Simulations also help us understand, why naturally occurring proteins fold with high precision into a unique three-dimensional structure, in which they can perform their biological function. Here we summarize recent results of a free-energy approach to simulate protein large-scale conformational change and folding with atomic precision. In the free-energy approach, which is based on Anfinsen's thermodynamic hypothesis, the conformational ensemble can be sampled with non-equilibrium methods, which accelerates the search of the high-dimensional protein landscape and permits the study of larger proteins at the all-atom level.

1 Introduction

Proteins are the workhorses of all cellular life. They constitute the building blocks and the machinery of all cells. Proteins perform a variety of roles in the cell: structural proteins constitute the building blocks for cells and tissues, enzymes, like pepsin, catalyze complex reactions, signaling proteins, like insulin, transfer signals between or within the cells. Transport proteins, like hemoglobin, carry small molecules or ions, while receptor proteins like rhodopsin generate response to stimuli. The mechanisms of all these biophysical processes depend on the precise folding of their respective polypeptide chains[1].

From the work of C.B. Anfinsen and co-workers in the 1960s we know that the amino acid sequence of a polypeptide chain in the appropriate physiological environment can fully determine its folding into a so-called native conformation[2]. Unlike man-made polymers of similar length, functional proteins assume unique three-dimensional structures under physiological conditions and there must be rules governing this sequence-to-structure

transition. Protein structures can be determined experimentally, by X-ray crystallography[3] or NMR methods[4], but these experiments are still challenging and do not work for all proteins. From the theoretical standpoint it is still not possible to reliably predict the native three-dimensional conformation of most proteins given their amino acid sequence alone[5–8].

The triplet genetic code by which the DNA sequence determines the amino acid sequence of polypeptide chains is well understood. However, unfolded polypeptide chains lack most of the properties needed for their biological function. The chain must fold into its native three dimensional conformation in order to perform its function[9]. Despite much research in this direction and the emergence of novel folding paradigms during the last decade, much of the mechanism by which the protein performs this auto-induced folding reaction is still unclear[6].

Therefore it would be very helpful to develop methods for protein structure prediction on the basis of the amino acid sequence alone. Even if this goal it is not fully realized, methods that can complete partially resolved experimental protein structures would be very helpful to determine the structure of proteins where neither theoretical methods nor experimental techniques alone can succeed[10]. For the trans-membrane family of proteins, present day experimental methods fail, which is responsible for the entire communication of the cell with its environment[11]. Theoretical methods would be very helpful to investigate these proteins. There are large number of related questions, for instance regarding the interactions of a given protein with a large variety of other proteins, where theoretical methods could also contribute to our understanding of biological function.

Related to the question of protein structure prediction is the question of how the proteins attain their final conformation - the so called protein folding problem. It remains one of the astonishing mysteries responsible for the evolution of life how these complex molecules can attain a unique native conformation with such precision. No man-made polymer of similar size is able to assemble into a predetermined structure with the precision encountered in the proteins that have evolved in nature.

Given its complexity it is not surprising that the protein folding process occasionally fails, and many of such failures are related to cellular disfunction or disease[12,13]. Therefore it is important not only to be able to predict the final structure of proteins but also very desirable to understand the mechanisms by which proteins fold.

Many theories and computational methods have been developed to understand the folding process. Simplified models have been applied to understand its physical principles[14]. Lattice based methods were among the first models that allowed efficient sampling of conformational space[15–17]. The lattice models, either 2D square or 3D cubic, were used to study protein folding and unfolding, but they were too simplified for protein structure prediction. Subsequently "Gō-Models" were developed, where only native contacts interact favorably[18], and were useful to characterize some aspects of the folding of small proteins. Further development led to statistically obtained knowledge based potentials[19–21]. These potentials were obtained and parameterized on the structures available from the Protein Data Bank. The knowledge based potentials are mostly used for fold recognition or protein structure prediction.

With the increase in computational resources and speed, all-atom molecular dynamics simulations of protein folding have been undertaken. For most proteins, it is still not feasible to determine the protein structure from extended conformations using a single molecular dynamics simulation. This is due to the fact that at the all-atom level, the typical

time step in a molecular dynamics simulation is about 1-2 femtoseconds while the protein folding occurs at millisecond timescale. A single such simulation would need years to complete. Replica exchange MD simulations have been successful in folding proteins from extended conformations, but are still limited to the size of 20-30 amino acids[22–26].

In this review we explore an alternative approach for protein structure prediction and folding that is based on the Anfinsen's hypothesis[2] that most proteins are in thermodynamic equilibrium with their environment in their native state. For proteins of this class the native conformation corresponds to the global optimum of the free energy of the protein. We know from many problems in physics and chemistry that the global optimum of a complex energy landscape can be obtained with high efficiency using stochastic optimization methods[27–29]. These methods map the folding process found in nature onto a fictitious dynamical process that explores the free-energy surface of the protein. By construction these fictitious dynamical processes not only find the conformation of lowest energy, but typically characterize the entire low-energy ensemble of competing metastable states.

This review is structured as follows: The second section introduces the protein the protein free-energy forcefield PFF02 and methods to efficiently explore the protein free-energy surface with stochastic simulation methods. In the next section, we review all-atom folding simulations for various proteins with the free-energy approach. The key results of these investigations and opportunities for further work are outlined in the last section.

2 Free-Energy Forcefields and Simulation Methods

2.1 The free-energy forcefield PFF02

We have recently developed an all-atom (with the exception of apolar CH_n groups) free-energy protein forcefield (PFF01) that models the low-energy conformations of proteins with minimal computational demand.[9,14] The forcefield parameterizes the internal free energy of a particular protein backbone conformation, excluding backbone entropy and thus makes different discrete conformational states directly comparable with regard to their stability. The effect of backbone entropy of a particular state can be assessed with Monte Carlo simulations at a finite temperature.

PFF02 contains the following non-bonded interactions:

$$V\left(\{\vec{r_i}\}\right) = \sum_{ij} V_{ij}\left[\left(\frac{R_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{ij}\frac{q_i q_j}{\varepsilon_{g(i)g(j)} r_{ij}}$$

$$+ \sum_i \sigma_i A_i + \sum_{hbonds} V_{hb} + V_{tor}$$

Here r_{ij} denotes the distance between atoms i and j and g(i) the type of the amino acid i. The Lennard Jones parameters (V_{ij},R_{ij}) for potential depths and equilibrium distance) depend on the type of the atom pair and were adjusted to satisfy constraints derived from a set of 138 proteins of the PDB database.[18−20] The non-trivial electrostatic interactions in proteins are represented via group-specific and position dependent dielectric constants $\varepsilon_{g(i)g(j)}$, depending on the amino-acids to which the atoms i and j belong. Interactions with the solvent were first fit in a minimal solvent accessible surface model[21] parameterized by free energies per unit area σ_j to reproduce the enthalpies of solvation of the

Gly-X-Gly family of peptides[22]. A_j corresponds to the area of atom i that is in contact with a fictitious solvent.

Hydrogen bonds are described via dipole-dipole interactions included in the electrostatic terms and an additional short range term for backbone-backbone hydrogen bonding (CO to NH) which depends on the OH distance, the angle between N, H and O along the bond and the angle between the CO and NH axis.[9] In comparison to PFF01, the forcefield PFF02 contains an additional term that differentiates between the backbone dipole alignments found in different secondary structure elements (included in the electrostatic potential between atoms i and j belonging to the backbone NH or CO groups via the dielectric constants $\varepsilon_{g(i)g(j)}$)[23] and a torsional potential for backbone dihedral angles V_{tor}, which gives a small contribution (about 0.3 kcal/mol) to stabilize conformations with dihedral angles in the beta sheet region of the Ramachandran plot.[14,24]

2.2 Stochastic Simulation Methods

Proteins assume unique three dimensional structures after being synthesized into a linear chain of amino acids. In the free-energy approach this native conformation corresponds to the global optimum of the free-energy forcefield. In order to fold proteins with free-energy methods, we need to use efficient sampling methods to reliably locate the associated global minima of the free-energy surface. The low-energy region of the free-energy landscape of proteins is extremely rugged due to the close packing of the atoms in the native conformation. Sampling this surface efficiently is therefore the central cogmputational bottleneck of this approach.

2.2.1 Monte Carlo

Most stochastic methods originate from the Monte Carlo method that explores the energy landscape by random changes in the geometry of the molecule. In this way large regions of the configurational space can be searched in finite time, without regard of the kinetics of the process. A Monte Carlo simulation is composed of the following steps:

1. Specify the initial coordinates (R_0).
2. Generate new coordinates by random change to initial coordinates ($R^{'}$).
3. Compute transition probability $T(R_0, R^{'})$.
4. Generate a uniform random number RAN in range [0,1].
5. If $T(R_0, R^{'}) < RAN$, then discard the new coordinates and goto step 2.
6. Otherwise accept the new conformation and goto step 2.

The most popular realization of the Monte Carlo method for molecular systems is the Metropolis method (see flowchart in Figure 1), which uses $T(R_0, R^{'}) = e^{-\Delta V/kT}$ if $\Delta V > 0$, and unit probability otherwise.

In Monte Carlo simulations, the system has no "memory" between two steps, *i.e.*, the probability that the system might revert to its previous state is as probable as choosing any

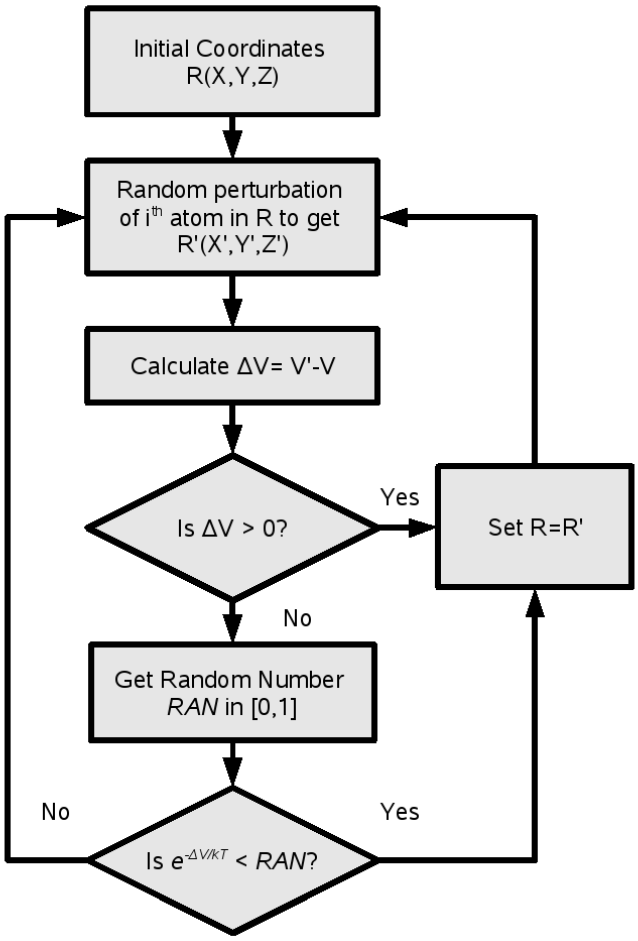


Figure 1. Schematic representation of Metropolis method.

other state. As a result of the stochastic simulation a large number of configurations is accumulated, which can be used to calculate thermodynamic properties of the system. Monte Carlo is not a deterministic method (as molecular dynamics), but gives rapid convergence of the thermodynamic properties[30].

2.2.2 Improved Sampling Techniques

Due to its popularity a large number of modifications and improvements of the Monte Carlo technique have been suggested and many of them have been used in the context of protein simulations:

- Simulated annealing: In this approach[31] barriers in the simulation are avoided by

starting the simulation at some high temperature and slowly lower the temperature of the simulation until the target temperature is reached. At high temperature the exploration of the phase space is very rapid, while near the end of the simulation the true thermodynamic probabilities of the system are sampled.

- Stochastic tunneling: Here a potential energy surface is transformed by using a non-linear transformation to suppress the barriers which are significantly above the present best energy estimate[32]. The transformed energy surface which is used for exploration of global minimum is given by

$$E_{STUN} = ln(x + \sqrt{x^2 + 1})$$

with $x = \gamma(E - E_0)$, where E is the present energy, E_0 is best estimation so far and γ the transformation parameter, which controls the rate of rise for the transformation.

- Parallel tempering: This method is Monte Carlo implementation of the replica exchange molecular dynamics method described. A modified version of this method, which uses an adaptive temperature control and replication step, has been employed for exploration of protein energy surfaces[33].

- Basin hopping technique (BHT): In this scheme the original potential energy surface is simplified by replacing the energy of each conformation with the energy of a nearby local minimum[34]. The minimization is carried out on the simplified potential (see section 2.2.3).

- Evolutionary strategy: This scheme is a multi-process extention of the BHT. Several concurrent simulations are carried out in parallel on a population. The population is evolved towards a global optimum of energy with a set of rules which enforce energy improvement and population diversity (see section 2.2.4).

2.2.3 Basin Hopping Technique

BHT[35] employs a relatively straightforward approach to eliminate high-energy transition states of the free-energy surface: The original free-energy surface is simplified by replacing the energy of each conformation with the energy of a nearby local minimum. In many applications the additional effort for the minimization step is more than compensated by the improved efficiency of the stochastic search. This process leads to a simplified potential on which the simulations search for the global minimum. This replacement eliminates high-energy barriers in the stochastic search that are responsible for the freezing problem in simulated annealing. A one dimensional schematic representation of BHT is shown in Figure 2. Every basin hopping cycle (minimization step) tries to locate a local minima and thus it simplifies the original potential energy surface (PES) (black curve) into an effective PES (blue curve) which is then searched for the global minima.

The basin hopping technique and its derivatives have been used previously to study the potential-energy surface of model proteins and polyalanines using all-atom models[36–39]. Here we replace the gradient-based minimization step used in many prior studies with a simulated annealing run[31], because local minimization generates only very small steps on the free energy surface of proteins. In addition, the computation of gradients for the SASA

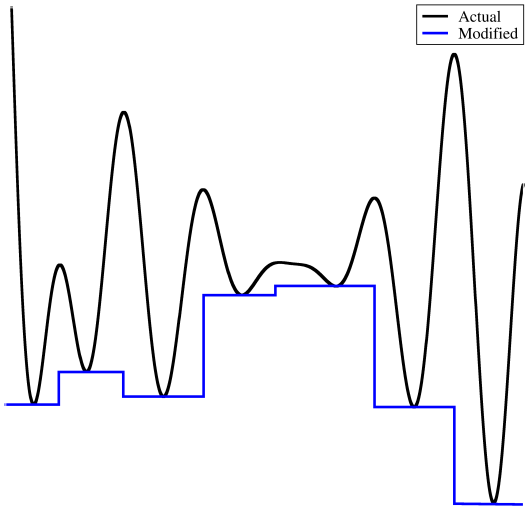


Figure 2. Schematic representation of Basin Hopping technique. The modified potential is obtained by replacing every point on the curve to its neared local minimum.

(Solvent Accessible Surface Area) is computationally prohibitive. Within each simulated annealing simulation, new configurations are accepted according to the Metropolis criterion, while the temperature is decreased geometrically from its starting to the final value.

The starting temperature and cycle length determine how far the annealing step can deviate from its starting conformation. The final temperature must be chosen small compared to typical energy differences between competing metastable conformations, to ensure convergence to a local minimum. The annealing protocol is thus parameterized by the starting temperature T_S, the final temperature T_F, and the number of steps. We investigated various choices for the numerical parameters of the method but have always used a geometric cooling schedule. At the end of one annealing cycle the new conformation is accepted if its energy difference to the current configuration was no higher than a given threshold energy ϵ_T, an approach recently proven optimal for certain optimization problems[40]. We typically used a threshold acceptance criteria of 1-3 kcal/mol.

2.2.4 Evolutionary Algorithms

The popular BHT method [41,34] for global optimization eliminates high-energy potential-energy surface (PES) by replacing the energy of each conformation with the energy of a nearby local minimum. For protein folding we have replaced the original local minimization by simulated annealing(SA). In the course of our folding studies, we find that independent BHT simulations often find identical structures corresponding to same local(global) minimum. As a result, each independent simulation reconstructs the full folding path independently. It would be very desirable to develop methods, where several concurrent simulations exchange information to *learn* from each other. For a PES having many local

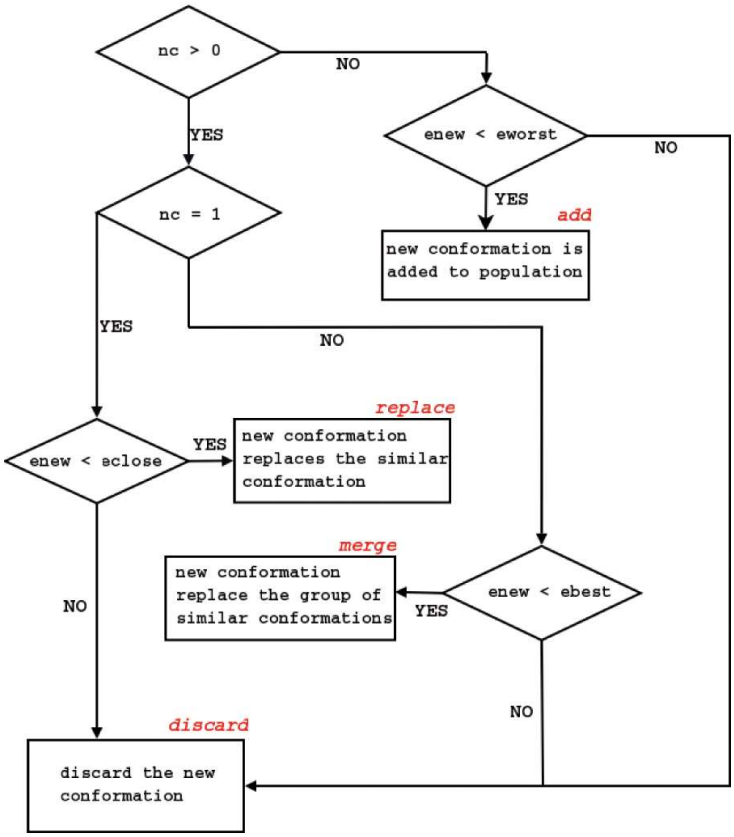


Figure 3. A flowchart illustrating the population update. See the text for an explanation

minima, independent simulations limit the efficient exploration of the PES. Also, occasionally BHT simulations go astray, ending the search in a wrong energy basin of the PES. We have developed a *greedy* version of BHT[42] which overcome these problems to a certain extent.

We have therefore generalized the BHT approach to a population of size N which is iteratively improved by P concurrent dynamical processes[33]. The population is evolved towards a optimum of the free energy surface with a ES that balances the energy improvement with population diversity. In the ES, conformations are drawn from the *active* population and subjected to an annealing cycle. At the end of each cycle the resulting conformation is either integrated into the active population or discarded. The algorithm was implemented as a master-client model in which idle clients request a task from the master. The master maintains the *active* conformation of the population and distributes the work to the clients. Each step in the algorithm has three phases:

1. Selection: A conformation is drawn randomly from the *active* population. We have used a uniform probability distribution with population of 20 conformers.

2. Annealing cycle: We use a simulated annealing schedule with T_{start} drawn from an exponential distribution and T_{end} fixed at 2K. The number of steps per cycle is increased as $10^5 \times \sqrt{cycle}$.

3. Population update: We have adjusted the acceptance criterion for newly generated conformations to balance the population diversity and energy enrichment. We define the two structures as *similar* if they have bRMSD less than 3 Å to each other. We define an *active* population as the pool containing mutually different lowest energy conformers. The master finds number of similar structures(*nc*) and then performs one of the following operations on complete population.

   (a) Add: If the new conformation is not *similar* to any structure(nc=0) in the population, we add it to the population, provided its energy is less than the energy of conformation with highest energy(E_{worst})

   (b) Replace: If the new conformation (with energy E_{new}) is *similar to one* existing structure in the population (with energy E_{old}), it replaces that structure provided $E_{\mathrm{new}} < E_{\mathrm{old}} + \Delta$ (see below).

   (c) Merge: If the new conformation has *several similar* structures, it replaces this group of structures provided its energy is less than the best one of the group E_{best} plus an acceptance threshold Δ.

   A flowchart illustrating the population update tasks of the master is shown in Fig. 3. In our first BHT/ES simulations we have used a fixed energy threshold (Δ) acceptance criterion. Here we have implemented a *variable* energy threshold which we define as $\Delta = A \times tanhD$, where

$$D = \frac{E_{\mathrm{new}} - E_{\mathrm{best}}}{A},$$

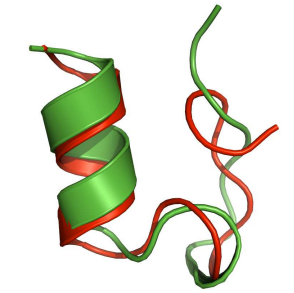
   where A is the energy threshold (3kcal/mol), Enew is energy of the new structure, E_{best} is the lowest energy structure in the population. This choice of the energy criterion ensures that the conformation with the best energy is never replaced, while conformations higher in energy are more easily replaced in the secure knowledge that they are far from optimal. The rules for the *replace* and *merge* operations ensure the structural diversity of the population and its continued energetic improvement (on average).

3 Folding Simulations

3.1 Helical Proteins

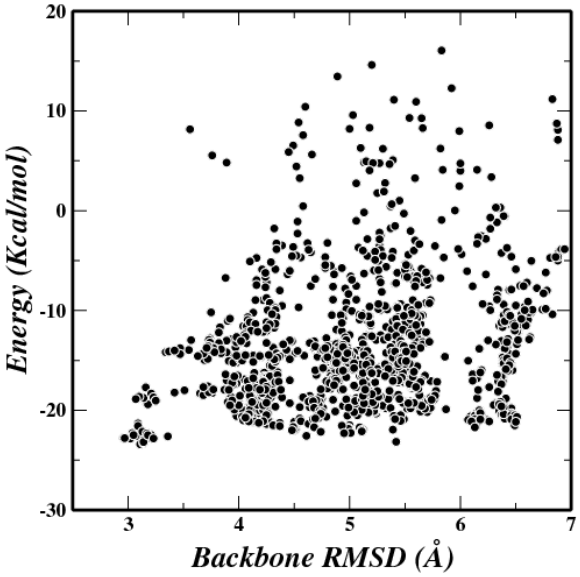
3.1.1 The tryptophan cage miniprotein

Tryptophan cage or trp-cage protein[43] has been the subject of various theoretical studies and it has been of great scientific interest. It had been reported to fold using replica



(a) Overlay

(b) C_β-C_β matrix



(c)

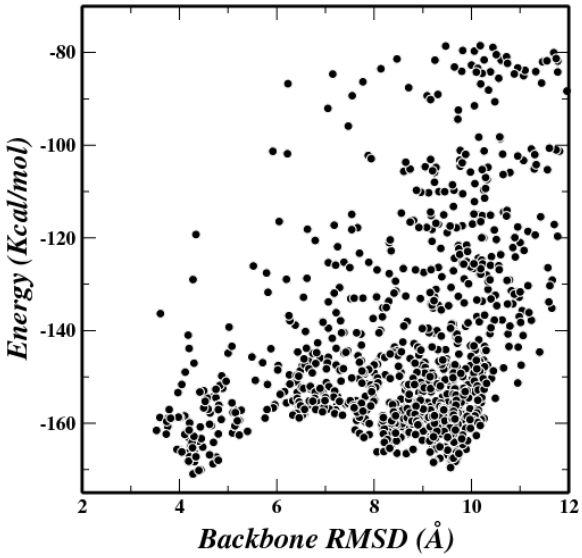
Figure 4. 1L2Y: Overlay of predicted (red) structure to experimental (green) structure. C_β-C_β distance overlay matrix and Energy vs. RMSD plot.

exchange MD and a variety of other simulations[44,27,45–47,29,48]. We performed 20 independent basin hopping simulations starting with the completely extended conformations in PFF02 with 100 cycles. The starting conformation had a RMSD of 12.94 Å to the native conformation and was completely extended manually (by setting all backbone dihedral angles except proline to 180°). The starting temperatures were chosen from a distribution of exponentially distributed temperatures and the number of steps increased with the BHT cooling cycle by $10^4\sqrt{n_m}$ where n_m is the number of minimization cycles.

The lowest energy structure converges to a native like conformation with RMSD of 3.11 Å to the native conformation. For the sake of uniformity in case of NMR resolved experimental structures, we compare the RMSD to the first model in the protein data bank file. The lowest energy structure had an energy of -23.4 Kcal/mol. Figure 4(c) shows the scatter plot of the conformations visited by the basin hopping simulations on the free energy surface. The overlay of native conformation (green) with the lowest energy conformation (red) is shown in Figure 4(a) and the corresponding C_β-C_β overlay matrix is shown in Figure 4(b). The C_β-C_β overlay matrix quantifies the tertiary alignment along with secondary structure formation by taking the difference between all C_β distances of predicted and native conformation. Black regions indicate excellent agreement in the formation of native contacts while white regions indicate larger deviations.

(a) Overlay

(b) C_β-C_β matrix



(c)

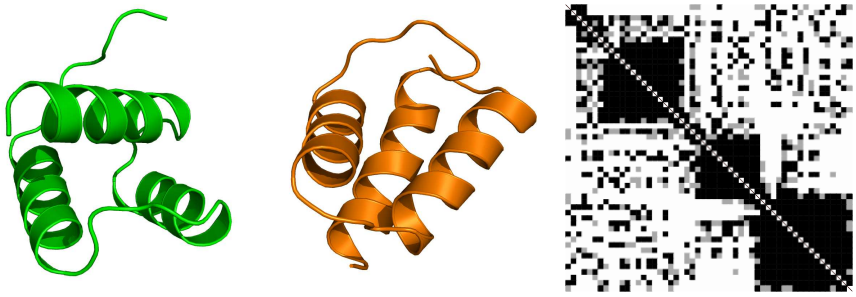
Figure 5. 1ENH: Overlay of predicted (red) structure to experimental (green) structure. C_β-C_β distance overlay matrix and Energy vs. RMSD plot.

3.1.2 The engrailed Homeodomain - 1ENH

The 54 amino acid engrailed homeodomain protein[49] is a three helical orthogonal bundle protein which has been subjected to detailed molecular dynamics simulations[50,51]. It was not possible to fold this protein using basin hopping technique due to the previously described freezing problem in the basin hopping simulations.

Here we studied the folding of engrailed homeodomain in PFF02 using the evolutionary algorithm with a maximum population of 64 conformations and 512 processors[52]. The lowest energy structure converges to 4.28 Å to the native conformation with the energy of -170.95 Kcal/mol. 1ENH has a unstructured tail at the N-terminus; after excluding this seven amino acid region, the RMSD reduces to only 3.4Å.

The scatter plot of conformations visited during the simulation are shown in Figure 5(c). Seven out of the total population of 64 structures are less than 4.5 Å RMSD to the native conformation. The overlay of the lowest energy conformation (red) with the native conformation (green) is shown in Figure 5(a) and the corresponding C_β-C_β overlay matrix is shown in Figure 5(b). There are also competing conformations (within 2 Kcal/-mol) with large RMS deviations encountered in the simulations. One such conformation is shown in Figure 6). These conformations have the same secondary structure, but a different tertiary structure alignment. The C_β-C_β overlay matrix for the misfolded conformation also confirms that all the three helices are properly predicted but their tertiary arrangement



(a) (b) (c)

Figure 6. 1ENH: Overlay of misfolded (orange) structure to experimental (green) structure and C_β-C_β distance overlay matrix.

is completely different. This indicates that various conformations exist in the low energy region of the 1ENH which are similar in secondary structure content.

No two helices in the misfolded conformation are in agreement with the respective helices in the native state. Independently, helix-1 (E8-E20), helix-2 (E26-L36) and helix-3 (A40-K43) are nearly perfectly predicted and have RMS of only 0.56, 0.42 and 0.47 Å respectively.

As about 10% of the population is native-like and the misfolded conformations we can conclude that the folding is reproducible.

3.2 Hairpins

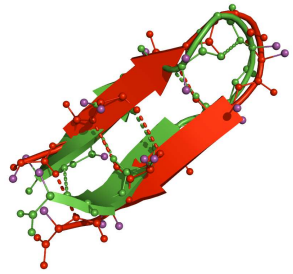
Hairpins are the simplest beta sheet structures with only two strands in antiparallel directions that are connected together with a turn. Hydrogen bonding and the packing of the protein itself plays a crucial role here in the folding of such small polypeptides. There are not many hairpin proteins that are not stabilized by external interaction with ions or with the formation of disulphide bridges.

3.2.1 trp-zippers

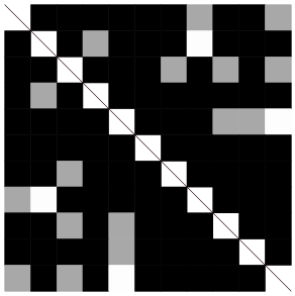
The tryptophan zippers are small monomeric stable β-hairpins that adopt an unique tertiary fold without requiring metal binding, unusual amino acids, or disulfide crosslinks[53]. We were able to fold various tryptophan zippers using PFF02 and basin hopping technique (not shown here).

We studied the folding of 1LE0 with EA using 128 processors on Marenostrum cluster at the Barcelona supercomputer center starting from completely extended conformations. We performed twenty cycles of evolutionary algorithm. The lowest energy conformation reached in the simulation had a RMSD of only 1.5 Å to the native conformation with the energy of -29.97 Kcal/mol.

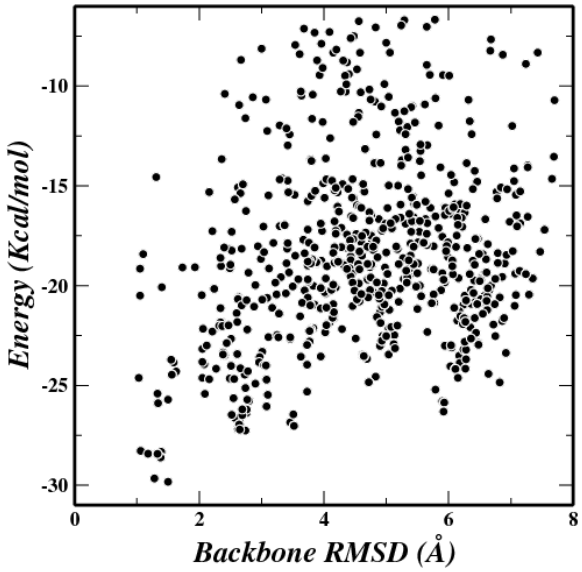
The scatter plot of the conformations visited during the simulations is shown in Figure 7(c). The scatter plot shows that the native-like conformations lie significantly below any other conformation. Twelve out of the 64 conformations from the final population



(a) Overlay



(b) C_β-C_β matrix



(c)

Figure 7. 1LE0: Overlay of predicted (red) structure to experimental (green) structure. C_β-C_β distance overlay matrix and Energy vs RMSD plot.

are less than 3.0 Å to the native conformation. The protein folds in less than 90 minutes using 128 processors in parallel by means of the twenty cycles of evolutionary algorithm amounting to 77×10^6 function evaluations or about 9 CPU days.

The overlay of the predicted conformation (red) with the native conformation (green) is shown in Figure 7(a) and the corresponding C_β-C_β overlay matrix is shown in Figure 7(c). Large black regions in the C_β-C_β overlay matrix indicates the agreement of native contacts between the two conformations.

As hydrogen bonding plays an important role in the formation and topology of β-sheet structures, it is important to compare the hydrogen bonding pattern in the lowest energy conformations as two β-sheet conformations might look very similar to the eye, but they might have completely different topology resulting from shifting of backbone hydrogen bonds.

The pattern of backbone hydrogen bonds is shown in Table 1 for the native and the predicted conformation. These were calculated with MOLMOL using the standard definitions (Distance=2.4Å and angle=35°). Four out of the five backbone hydrogen bonds of the native structure are predicted correctly in the lowest energy structure found in the simulations.

As about 20% of the population converged to native-like conformations with much lower energies, we conclude the folding of tryptophan zipper as reproducible and predictive.

Hydrogen bond	Native	Predicted
03 THR HN ⟶ 10 THR O	X	X
05 GLU HN ⟶ 08 LYS O	X	X
07 ASN HN ⟶ 05 GLU O	X	
10 THR HN ⟶ 03 THR O	X	X
12 LYS HN ⟶ 01 SER O	X	X

	Secondary Structure	RMSD (Å)
Native	CEEECSSSEEEC	-
Predicted	CEEEETTEEEEC	1.52

Table 1. 1LE0: Backbone hydrogen bond pattern of the native and predicted conformations and secondary structure information.

3.2.2 HIV-1 V3 loops

We studied the folding of 14 amino acid HIV-1 $V3_{MN}$ loop 1NIZ[54] in PFF02 using a greedy version of the basin hopping technique[55].

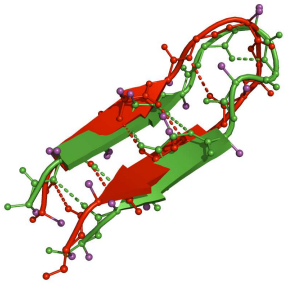
In basin hopping simulations there is a threshold energy acceptance criterion at the end of every basin hopping cycle. In our previous simulations, we have used this threshold acceptance criterion of 1-3 Kcal/mol depending upon this size of the protein. In the greedy version of basin hopping the threshold energy is varied depending upon the best energy found so far in the simulation. Here we calculated the threshold as $(\epsilon_S - \epsilon_B)/4$, where ϵ_S is the starting energy and ϵ_B is the best energy found so far in the simulation. This choice implies that the conformation with the best energy is never replaced with a conformation that is higher in energy and thus introduces a "memory effect" in the simulation. For the simulations that are higher in energy, the increased threshold value implies a higher acceptance probability of conformations with higher energy.

We did 200 cycles of greedy basin hopping simulations in PFF02. The simulations were started with completely extended conformation that had the RMSD of 12 Å to the native state. The lowest energy structure found in the simulation had the RMSD of only 2.04 Å to the native state.

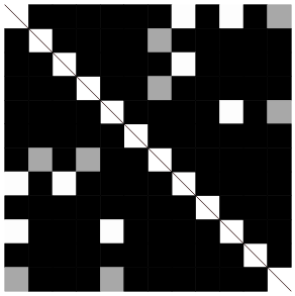
Hydrogen bond	Native	Predicted
02 ARG HN ⟶ 13 THR O	X	X
04 HIS HN ⟶ 11 PHE O	X	X
06 GLY HN ⟶ 09 ARG O		X
08 GLY HN ⟶ 06 GLY O	X	
11 PHE HN ⟶ 03 HIS O	X	X
13 THR HN ⟶ 01 ARG O	X	X

	Secondary Structure	RMSD (Å)
native	CEEEECSSCEEEEC	-
predicted	CEEEECSSCEEEEC	2.04

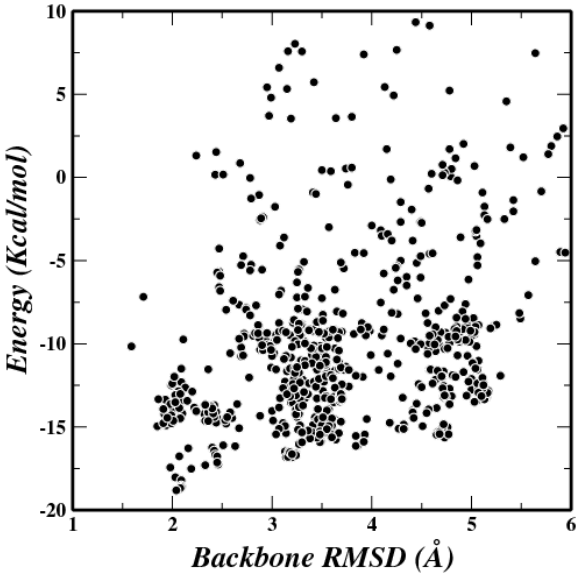
Table 2. 1NIZ: Backbone hydrogen bond pattern between native and predicted conformations and secondary structure information.



(a) Overlay



(b) C_β-C_β matrix



(c)

Figure 8. 1NIZ: Overlay of predicted (red) structure to experimental (green) structure. C_β-C_β distance overlay matrix and Energy vs. RMSD plot.

The scatter plot of the conformations visited during the simulations is shown in Figure 8(c). The scatter plot shows a single downhill folding funnel for this hairpin. Eight out of the ten independent simulations converged to less than 3.5 Å RMSD to the native conformation.

The overlay of the lowest energy conformation (red) with the native conformation (green) is shown in Figure 8(a) and the corresponding C_β-C_β distance matrix is shown in Figure 8(c). Large black regions in the C_β-C_β overlay matrix indicates the agreement of native contacts between the two conformations.

Again, we did the backbone hydrogen bond analysis. Four out of the five backbone hydrogen bonds of the native structure were correctly predicted in the lowest energy structure found in the simulations. The pattern of the backbone hydrogen bonds is shown in Table 2. The secondary structure of the predicted and native conformation is also shown in Table 2. The letters in the secondary structure correspond to DSSP definitions.

As eight of the ten simulations converged to the native-like conformation without any competing metastable conformations, the folding is concluded as reproducible and predictive.

3.3 A mixed secondary structure protein

Zinc fingers are among the most abundant proteins in eukaryotic genomes and occur in many DNA binding domains and transcription factors[56]. They participate in DNA recognition, RNA packaging, transcriptional activation protein folding and assembly and apoptosis. Many zinc fingers contain a $\mathrm{Cys_2His_2}$ binding motif that coordinates the Zn-ion in $\alpha\beta\beta$ -framework[57–59] and much effort is towards the engineering of novel zinc fingers[60]. A classical zinc finger motif binding DNA is illustrated in Fig. 9.

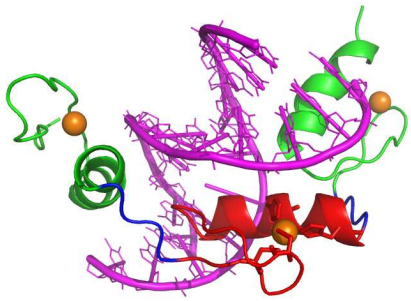


Figure 9. A classical $\mathrm{Cys_2His_2}$ zinc finger motif with Zn-ion(orange) and DNA (magenta).

The reproducible folding of such proteins with mixed secondary structure, however, remains a significant challenge to the accuracy of the all-atom forcefield and the simulation method[61]. We use the all-atom free-energy forcefield PFF02 to predictively fold the 23-51 amino-acid segment of the N-terminal sub-domain of ATF-2 (PDBID 1BHI)[62], a 29 amino acid peptide that contains the basic leucine zipper motif . 1BHI folds into the classical TFIIIa conformation found in many zinc-finger like sub-domains. The fragment contains all the conserved hydrophobic residues (PHE25, PHE36, LEU42) of the classical zinc finger motif and the CYS27, CYS32, HIS45, HIS49 zinc binding pattern.

Starting from a completely unfolded conformation with no secondary structure (16 Å backbone RMSD (bRMSD) to native) we performed 200 cycles of the evolutionary algorithm. The distribution of bRMSD versus energy of all accepted conformations during the simulation (Fig. 10) demonstrates that the simulation explores a wide variety of conformations, with regard to their free-energy and their deviation from the native conformation.

Among the ten energetically lowest conformations (see Table 3) six fold into near-native conformations with bRMSDs of 3.68-4.28 Å, while four fold to conformations with a larger bRMSD. The three energetically best conformations are all near-native in character. An overlay with the experimental conformation (left panel of Fig. 11) illustrates that the helix, beta-sheet and both turns are correctly formed. The hydrophobic residues, which determine the packing of the beta-sheet against the helix, are illustrated in blue in the figure. The helical section (GLU39-GLU50) and the beta-sheet (PHE25-LEU26 and ARG35-PHE36) deviate individually by 1.6 Å and 2.4 Å bRMSD from their experimental counterparts, respectively.

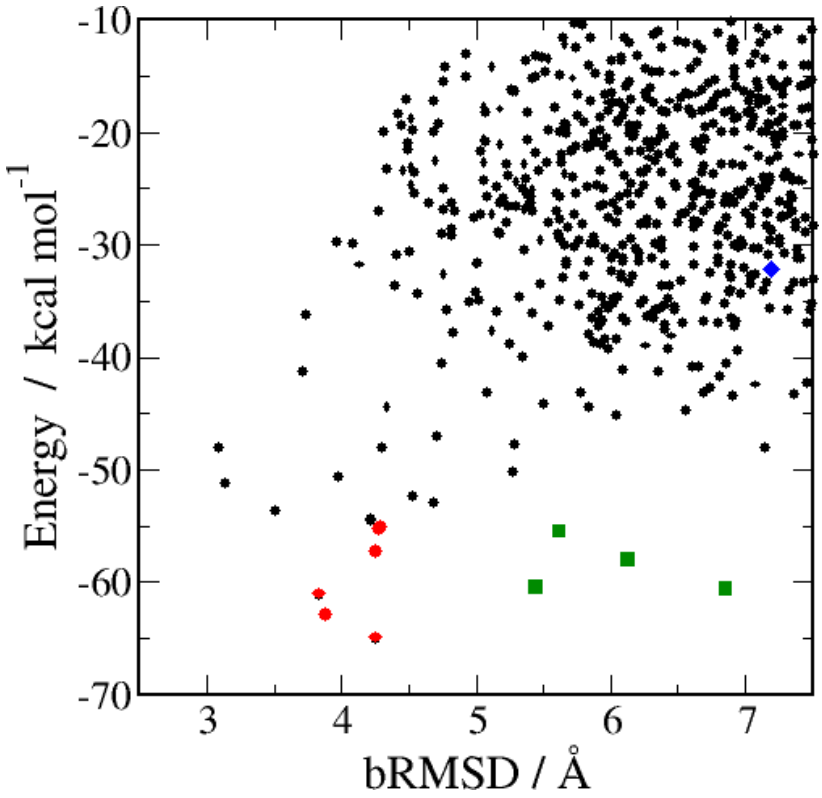


Figure 10. Free energy versus bRMSD of all accepted conformations in the simulation. The best 10 structures are highlighted as: red circles(native-like), green squares(non-native). The folding intermediate is denoted by blue diamond

The overall difference between the experimental and the folded conformations stems from the relative arrangement of the beta-sheet with respect to the helix, which is dominated by unspecific hydrophobic interactions. All conserved hydrophobic sidechains are also buried in the folded structure. The zinc-coordinating cysteine residues (CYS27,CYS32) are within 2 Å of their native positions and available association with the Zn-ion.

Fig. 12 shows the convergence of the energy. After about 120 attempted updates per population member (3.5×10^8 function evaluations) the population converged to the native ensemble. According to the funnel paradigm for protein folding[63], tertiary structure forms as the protein slides downhill on the free-energy surface from the unfolded ensemble towards the native conformation. Each annealing cycle generates a small perturbation on the existing conformation, which averages to a 0.5 Å bRMSD change (max 3 Å initially). As new low-energy conformations replace old conformations, the population slides as a whole down the funnel of the free energy landscape.

Ensemble averages as a function of time over the moving population are thus associated with different stages of the structure formation process. In the lower panels of Fig. 12, we

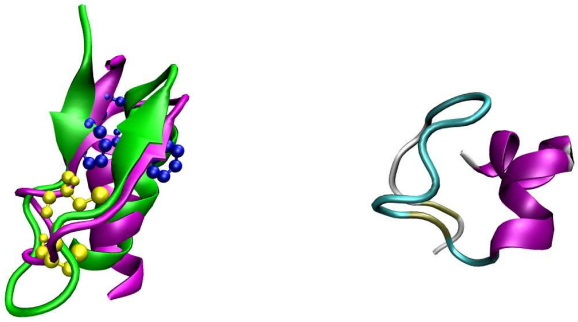


Figure 11. Left: Overlay of the native (green) and folded (magenta) conformations. The conserved hydrophobic residues are shown in blue and Zn binding cysteines are shown in yellow. Right: The intermediate conformation with partially formed helix and β sheet.

plot the average helical content and the number of beta-sheet H-bonds as a function of the cycle number. Following a rapid collapse to a compact conformation, the helix forms first, followed by the formation of the beta sheet. The analysis of the folding funnel upwards in energy illustrates that the lowest energy metastable conformations correspond to a partial unzipping of amino acids PHE25-ARG35, while the conserved cysteine residues are still buried. Even much higher on the free energy funnel (blue diamond in Fig. 10), we find many structures that have much residual structure, but essentially not long-range native contacts.

The preformed sheet-region is stabilized by the hydrogen bonds (LEU26-CYS27, ARG35) and packs at the right angle to the helix, the hydrophobic residues are only partially buried. This conformational freedom may be relevant in DNA binding, where the helical part of the zinc finger packs into the major groove of the DNA.

De novo folding of the zinc finger domain permits a direct sampling of the relevant low-energy portion of the free-energy surface of the molecule as the first step towards the elucidation of the structural mechanisms involved in DNA binding[64]. We find that much of the structure of the zinc finger is formed even in the absence of the metal ion that is ultimately required for the stabilization of the native conformation. Because the algorithm tracks the development of the population it is possible to reconstruct a folding pathway by reconstructing the sequence of events starting with converged conformation and moving backwards to the completely unfolded conformation.

We have thus demonstrated predictive all-atom folding of the DNA binding zinc-finger motif in the free-energy forcefield PFF02. This investigation offers the first unbiased characterization of the low-energy part of the free-energy surface of the zinc finger motif, which is unattainable in coarse grained, knowledge-based models. We find that the helix forms first along the folding path and acts as a template against which a variety of near-native beta-sheet backbone arrangements can pack. There are many zinc fingers with bRMSD

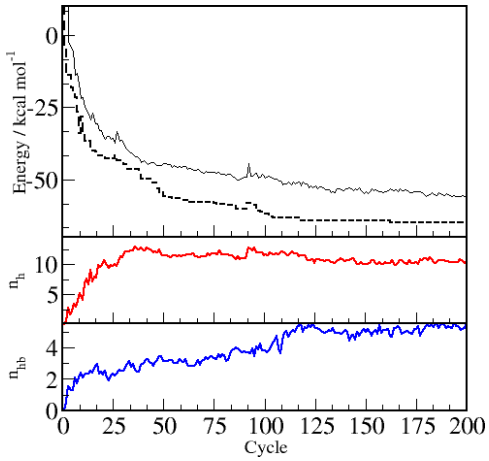


Figure 12. Top: Average (solid line) and best (dashed line) energies as functions of the number of the simulation cycle for the Zinc Finger, Middle: number of amino acids (n_h) in a helical conformation (as computed by DSSP) and Bottom: number of hydrogen bones (n_{hb}) as function of the ES cycle number

#	Energy kcal/mol	bRMSD Å	Secondary Structure
E01	-64.94	4.25	CC**EE**CTTTTSCC**EE**SSC**HHHHHHHHHHHH**C
E02	-62.84	3.88	CC**EE**CTTTTSCC**EE**SSC**HHHHHHHHH**STTC
E03	-61.05	3.83	CC**EE**CTTTTCCC**EE**SSC**HHHHHHHHH**STTC
E04	-60.51	6.85	CC**EE**CTTTTSCC**EE**CSC**HHHHHHH**SCCCCC
E05	-60.40	5.44	CCBBCTTTTCCCBCCSC**HHHHHHHH**CCCBC
E06	-57.93	6.12	CC**EE**CTTTTSCC**EE**CSC**HHHHHHH**SCCCCC
E07	-56.21	4.25	CC**EEEE**CSSSSC**EEEE**SC**HHHHHHHHHHH**C
E08	-55.44	5.61	CCSSSCSSCCSSCCCSC**HHHHHHHHH**TTTC
E09	-55.18	4.27	CCCC**EE**CTTSSC**EE**CS**HHHHHHHHHH**CSCC
E10	-55.02	-4.29	CCCCBTTTTBTTCCCSS**HHHHHHHHHHHH**C

Table 3. Energy, bRMSD and secondary structures of the 10 lowest energy structures

of less than 2 Å to 1BHI[62]. Thus, this investigation provides one important step in the theoretical understanding of zinc-finger formation and function.

4 Summary

These investigations demonstrate that the free-energy approach is able to predict the native state of a wide range of proteins at the global minimum of their free energy surface[27, 65–71]. Protein folding with free energy methods is much faster than the direct simulation of the folding pathway by kinetic methods such as molecular dynamics. Using just standard PCs we can fold a simple hairpin with fifteen to twenty amino acids in a matter of hours, at most in a day[69]. Unfortunately even for free energy methods the computational cost rises steeply with the system size.

The second ingredient in protein folding studies, aside from the force field, are therefore the simulation protocols, which ultimately determine whether the global optimum of the forcefield is determined accurately and reliably. We have reviewed key aspects of such methods, e.g. the stochastic tunneling or the basin hopping technique, which had proven successful in folding studies for small proteins. One of the key limitations of these methods is that they map the global optimization problem onto a single fictitious dynamical process, while in principle, many concurrent processes can be used[28, 29, 65].

We have therefore also discussed an evolutionary algorithm[72] for massively parallel architectures, such as the BlueGene architecture, which keeps a diverse population on the master, while the clients sample the protein landscape simultaneously. This algorithm scales very well with the number of processors used (up to 4096 tested on the IBM BlueGene). Using this algorithm we folded various proteins such as 40 amino acid HIV accessory protein (1F4I) and 54 amino acid engrailed homeodomain protein (1ENH) in a single day. The folding of the engrailed homeodomain protein was carried out in a single day using 512 processors on the Barcelona Mare Nostrum Supercomputer, the current largest supercomputer in Europe. Folding of the tryptophan zipper protein (1LE0) was possible in only 14 minutes using 128 processors[69].

To date we have succeeded to develop methods to find the native state of various proteins by locating the global minimum of the free energy surface[28]. There are, however, a large number of questions that remain to be addressed. Fortunately there are complementary methods, which in combination with the free-energy methodology developed here, can address these problems. For example, we have neglected the details of the kinetics of protein folding in our approach. As stated earlier, its important to study kinetics of folding to understand protein folding mechanism and to predict folding rates. Because free-energy methods sample exhaustively the low-energy conformations of the protein that are accessible under physiological conditions it may be possible to reconstruct the folding kinetics on the basis of that ensemble of conformations. This can be achieved by a dynamical analysis of the low energy region by using master equations assuming diffusive processes between similar conformations.

With the development of the all-atom protein forcefield (PFF02) we have made a significant step towards a universal free-energy approach to protein folding and structure prediction[68]. The massively parallel simulation methods developed in the last few years now permit the protein folding of medium-size proteins from random initial conformations. This work thus lays the foundations to further explore the mechanism of protein folding, to understand protein stability and ultimately develop methods for *de novo* protein structure prediction.

References

1. C. Branden and J. Tooze. *Introduction to protein structure.* Routledge, 1999.
2. C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.
3. L. Stryer. Implications of x-ray crystallographic studies of protein structure. *Annu. Rev. Biochem.*, 37:25–50, 1968.
4. G. Wagner, S. G. Hyberts, and T. F. Havel. Nmr structure determination in solution: A critique and comparison with x-ray crystallography. *Annu. Rev. Biophys. Biomol. Struct.*, 21:167–98, 1992.
5. Richard Bonneau and David Baker. Ab initio protein structure prediction: Progress and prospects. *Annu. Rev. Biophys. Biomol. Struct.*, 30:173–89, 2001.
6. D. Baker and A. Sali. Protein structure prediction and structural genomics. *Science*, 294:93–96, 2001.
7. J. Moult, K. Fidelis, A. Zemia, and T. Hubbard. Critical assessment of methods of protein structure (casp): round iv. *PROTEINS:Structure, Function, and Bioinformatics*, 45:2–7, 2001.
8. C. Hardin, M.P. Eastwood, M. Prentiss, Z. Luthey-Schulten, and P. Wolynes. Folding funnels: The key to robust protein structure prediction. *J. Comp. Chem.*, 23:138–146, 2003.
9. Jeremy M. Berg, John L. Tymoczky, and Lubert Stryer. *Biochemistry, fifth edition*. Michelle Julet, 2002.
10. B. Rost. Protein secondary structure prediction continues to rise. *J. Struct. Biol*, 134:204–18, 2001.
11. W. Khlbrandt and E. Gouaux. Membrane proteins. *Current Opinion in Structural Biology*, 9:445–7, 1999.
12. C. M. Dobson. The structural basis of protein folding and its links with human disease. *Phil. Trans. R. Soc. Lond. B*, 356:133–145, 2001.
13. M.B. Pepys. In J.G. Ledingham D.J. Weatherall and D.A. Warrel, editors, *The Oxford Textbook of Medicine (third ed.)*. Oxford University Press, Oxford, 1995.
14. H. S. Chan and K. A. Dill. Protein folding in the landscape perspective: Chevron plots and non-arrhenius kinetics. *Proteins: Struc. Func. and Gen.*, 30:2–33, 1998.
15. K. F. Lau and K. A. Dill. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules*, 22:3986–97, 1989.
16. Ken A. Dill, Sarina Bromberg, Kaizhi Yue, Klaus M. Fiebig, David P. Yee, Paul D. Thomas, and Hue Sun Chan. Principles of protein folding- a perspective from simple exact models. *Protein Science*, 4:561–602, 1995.
17. Eugene Shakhnovich, G. Farztdinov, A.M. Gutin, and Martin Karplus. Protein folding bottlenecks: A lattice monte carlo simulation. *Phys. Rev. Lett.*, 67(12):1665–1668, 1991.
18. N. Go and H. A. Scheraga. Analysis of the contribution of internal vibrations to the statistical weight of equilibirium conformations of macromolecules. *J. Chem. Phys.*, 51:4751–4767, 1969.
19. M. J. Sippl, G. Nemethy, and H. A. Scheraga. Intermolecular potentials from crstal data. 6. determination ofemperical potentials for o-h$\cdots$o=c hydrogen bonds from packing configurations. *J. Phys. Chem.*, 88:6231–6233, 1984.

20. G. Casari and M. J. Sippl. Structure derived hydrophobic potentials. a hydrophobic potential derived from x ray structures of globular proteins is able to identify native folds. *J. Molec. Biol.*, 224:725–732, 1992.
21. M. J. Sippl. Knowledge-based potentials for proteins. *Current Opinion in Structural Biology*, 5:229–35, 1995.
22. F. Rao and A. Caflisch. Replica exchange molecular dynamics simulations of reversible folding. *J. Chem. Phys.*, 119:4035–4042, 2003.
23. P. H. Nguyen, G. Stock E. Mittag, C. K. Hu, and M. S. Li. Free energy landscape and folding mechanism of a β-hairpin in explicit water: A replica exchange molecular dynamics study. *Proteins: Struc. Func. and Gen.*, 61:705–808, 2005.
24. Y.M. Rhee and V.S. Pande. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J*, 84:775–786, 2003.
25. S.-Y. Kim, J. Lee, and J. Lee. Folding of small proteins using a single continous potential. 120:8271–8276, 2004.
26. J.-E. Shea and C. L. Brooks III. From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.*, 52:499–535, 2001.
27. A. Schug, T. Herges, and W. Wenzel. Reproducible protein folding with the stochastic tunneling method. *Phys. Rev. Lett.*, 91:1581021–4, 2003.
28. A. Schug, B. Fischer, A. Verma, H. Merlitz, W. Wenzel, and G. Schoen. Biomolecular structure prediction stochastic optimization methods. *Advanced Engineering Materials*, 7(11):1005–1009, 2005.
29. A. Schug, T. Herges, A. Verma, K. H. Lee, and W. Wenzel. Comparison of stochastic optimization methods for all-atom folding of the trp-cage protein. *Chemphyschem*, 6:2640–6, 2006.
30. A. R. Leach. *Molecular Modelling: Principles and Applications*. Pearson Education Ltd., 2001.
31. S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–80, 1983.
32. K. Hamacher and W. Wenzel. A stochastic tunnelling approach for global minimization. *Phys. Rev. E*, 59:938, 1999.
33. A. Schug, T. Herges, A. Verma, and W. Wenzel. Investigation of the parallel tempering method for protein folding. *Phys. Cond. Matter, special issue: Structure and Function of Biomolecules (in press)*, 2005.
34. D. M. Leitner, C. Chakravarty, R. J. Hinde, and D. J. Wales. Global optimization by basin-hopping and the lowest energy structures of lennard jones clusters containing upto 110 atoms. *Phys. Rev E*, 56:363, 1997.
35. A. Nayeem, J. Vila, and H. A. Scheraga. A comparative study of the simulated-annealing and monte carlo-with-minimization approaches to the minimum-energy structures of polypeptides: [Met]-enkephalin. *J. Comp. Chem.*, 12:594–605, 1991.
36. R. A. Abagyan and M. Totrov. Biased probability monte carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.*, 235:983–1002, 1994.
37. D. J. Wales and P. E. J. Dewsbury. Effect of salt bridges on the energy landscape of a model protein. *J. Chem. Phys.*, 121:10284–90, 2004.
38. P. N. Mortenson and D. J. Wales. Energy landscapes, global optimisation and dynam-

ics of the polyalanine Ac(ala)8 NHMe. *J. Chem. Phys.*, 114:6443–54, 2001.
39. P. N. Mortenson, D. A. Evans, and D. J. Wales. Energy landscapes of model polyalanines. *J. Chem. Phys.*, 117:1363–76, 2002.
40. J. Schneider, I. Morgenstern, and J. M. Singer. Bouncing towards the optimum: Improving the results of monte carlo optimization algorithms. *Phys. Rev. E*, 58:5085–95, 1998.
41. A. Nayeem, J. Vila, and H.A. Scheraga. A comparative study of the simulated-annealing and monte carlo-with-minimization approaches to the minimum-energy structures of polypeptides: [met]-enkephalin. *J. Comp. Chem.*, 12(5):594–605, 1991.
42. W. Wenzel. Predictive folding of a β hairpin in an all-atom free-energy model. *Europhys. Letters*, 76:156, 2006.
43. J. W. Neidigh, R. M. Fesinmeyer, and N. H. Andersen. Designing a 20-residue protein. *Nat. Struct. Biol.*, 9:425–30, 2002.
44. C. D. Snow, B. Zagrovic, and V. S. Pande. Folding kinetics and unfolded state topology via molecular dynamics simulations. *J. Am. Chem. Soc.*, 124:14548–14549, 2002.
45. F. Ding, S. V. Buldyrev, and N. V. Dokholyan. Folding trp-cage to nmr resolution native structure using a coarse-grained protein model. *Biophys. J.*, 88:147–55, 2005.
46. A. Linhananta, J. Boer, and I. MacKay. The equilibrium properties and folding kinetics of an all-atom go- model of the trp-cage. *J. Chem. Phys*, 122:1–15, 2005.
47. A. Schug, W. Wenzel, and U. H. E. Hansmann. nergy landscape paving simulations of the trp-cage protein. *J. Chem Phys.*, 122:1–7, 2005.
48. J. Juraszek and P. G. Bolhuis. Sampling the multiple folding mechanisms of trp-cage in explicit solvent. *Proc. Natl. Acad. Sci. USA*, 103:15859–64, 2006.
49. N. D. Clarke, C. R. Kissinger, J. Desjarlais, G. L. Gilliland, and C. O. Pabo. Structural studies of the engrailed homeodomain. *Protein Sci.*, 3:1779–87, 1994.
50. U. Mayor, N. R. Guydosh, C. M. Johnson, J. G. Grossmann, S. Sato S, G. S. Jas, S. M. Freund, D. O. Alonso, V. Daggett, and A. R. Fersht. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature*, 421:863–7, 2003.
51. V. Daggett and A. Fersht. The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell. Biol.*, 4:497–502, 2003.
52. A. Verma and W. Wenzel. All-atom protein folding in a single day. submitted, 2006.
53. A. G. Cochran, N. J. Skelton, and M. A. Starovasnik. Tryptophan zippers: stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci. USA*, 98:5578–83, 2001.
54. M. Sharon, N. Kessler, R. Levy, S. Zolla-Pazner, M. Gorlach, and J. Anglister. Alternative conformations of HIV-1 V3 loops mimic β-hairpins in chemokines, suggesting a mechanism for coreceptor selectivity. *Structure*, 11:225–236, 2003.
55. A. Verma and W. Wenzel. De-novo all atom folding of a HIV-1 V3 hairpin loop in an imporved free energy forcefield. Submitted, 2006.
56. J. H. Laity, B. M. Lee, and P. E. Wright. Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol*, 11:39–46, 2001.
57. MS Lee, GP Gippert, KV Soman, DA Case, and PE Wright. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science*, 245(4918):635–637, 1989.
58. NP Pavletich and CO Pabo. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science*, 252(5007):809–817, 1991.

59. Scot A. Wolfe, Lena Nekludova, and Carl O. Pabo. Dna recognition by cys2his2 zinc finger proteins. *Annual Review of Biophysics and Biomolecular Structure*, 29(1):183–212, 2000.
60. Fyodor D. Urnov, Jeffrey C. Miller, Ya-Li Lee, Christian M. Beausejour, Jeremy M. Rock, Sheldon Augustus, Andrew C. Jamieson, Matthew H. Porteus, Philip D. Gregory, and Michael C. Holmes. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, 435(7042):646–651, June 2005.
61. A. Abagyan and M. Totrov. Ab initio folding of peptides by the optimal-bias monte carlo minimization procedure. *J Comput Phys*, 402-412:151, 1999.
62. A. Nagadoi, K. Nakazawa, H. Uda, K. Okuno, T. Maekawa, S. Ishii, and Y. Nishimura. Solution structure of the transactivation domain of atf-2 comprising a zinc finger-like subdomain and a flexible subdomain. *J. Mol. Biol.*, 287:593–607, 1999.
63. J. N. Onuchic, Z. Luthey-Schulten, and P.G. Wolynes. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.*, 48:545–600, 1997.
64. J. H. Laity, H. J. Dyson, and P. E. Wright. Dna-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in $cys_2 - his_2$ zinc fingers. *J. Mol. Biol*, 295:719–727, 2000.
65. A. Verma, A. Schug, K. H. Lee, and W. Wenzel. Basin hopping simulations for all-atom protein folding. *J. Chem. Phys.*, 124:044515, 2006.
66. A. Quintilla, E. Starikov, and W. Wenzel. De novo folding of two-helix potassium channel blockers. *J. Chem. Theory and Computation.*, 3:1183–92, 2007.
67. A. Schug, T. Herges, and W. Wenzel. All atom folding of the three helix hiv accessory protein with an adaptive parallel tempering method. *Proteins*, 57(4):792–798, 2004.
68. A. Verma and W. Wenzel. Towards an all-atom free-energy forcefield for protein folding. in preparation, 2006.
69. Abhinav Verma, Srinivasa M. Gopal, Jung S. Oh, Kyu H. Lee, and Wolfgang Wenzel. All-atom de novo protein folding with a scalable evolutionary algorithm. *J. Comp. Chem*, 28:2552–2558, 2007.
70. Srinivasa M. Gopal and Wolfgang Wenzel. De novo folding of the dna-binding atf-2 zinc finger motif in an all-atom free-energy forcefield. *Angewandte Chemie International Edition*, 45(46):7726–7728, 2006.
71. A. Verma and W. Wenzel. Protein structure prediction by all-atom free-energy refinement. *BMC Structural Biology*, 7:12, 2007.
72. A. Schug and W. Wenzel. Reproducible folding of a four helix protein in an all-atom forcefield. *J. Am. Chem. Soc.*, 126(51):16736–16737, 2004.

Multiscale Methods for the Description of Chemical Events in Biological Systems

Marcus Elstner[1,2] **and Qiang Cui**[3]

[1] Department of Physical and Theoretical Chemistry
Technische Universität Braunschweig
D-38106 Braunschweig, Germany
[2] Department of Molecular Biophysics
German Cancer Research Center
D-69115 Heidelberg, Germany
E-mail: m.elstner@tu-bs.de

[3] Department of Chemistry and Theoretical Chemistry Institute
University of Wisconsin-Madison
1101 University Avenue, Madison, WI 53706, USA
E-mail: cui@chem.wisc.edu

Computational methods for the description of chemical events in biological structures have to take into account the key features of bio-molecular molecules, their high degree of structural flexibility and the long-range nature of electrostatic forces. In the last decade, a multitude of approaches have been developed to combine computational methods that span different length- and time-scales. These multiscale approaches incorporate a quantum mechanical description of the actives site in combination with an empirical force field method for the immediate protein environment and a continuum treatment of the regions further away. To study reactive events, efficient sampling techniques have to be applied, which can become computationally intense and therefore requires effective quantum methods. In this contribution, we describe the various options to combine different methods, where the specific combination depends very much on the nature of the problem in hand.

1 Introduction

The simulation of structure and dynamics of biological systems can nowadays be routinely performed using empirical force fields, which have become robust and reliable tools over the last decades[1,2]. These Molecular Mechanics (MM) force fields[3,4] model chemical bonds by harmonic springs, i.e. they describe the energy of a chemical bond using harmonic (or Fourier) potentials for the bond length, bond angle and dihedral angle. In addition to these bonded terms, the force fields contain non-bonded contributions, modeled by the interaction of fixed atomic point charges and van der Waals interactions, usually described by the 12-6 Lennard-Jones potential. Polarizable force fields[5] that allow the partial charges to vary depending on their environment have also been developed, although their applications have been much more limited due to the higher computational expense.

Biological structures host a multitude of chemical events like chemical reactions (biocatalysis), photochemical processes, long range proton transfers (e.g., in bioenergetics), electron and energy (excitation) transfers, which can only be described using quantum mechanical (QM) techniques and not with MM. The description of these processes is very challenging for computational chemistry due to the large size of biological systems and the presence of multiple time-scales. Indeed, biological structures take the middle ground

between solids and more disordered materials like polymers. On the one hand, they have a highly ordered structure from a functional perspective; e.g., specific functional amino acids with pre-organized orientations are found in the immediate vicinity of the active site, which is one important reason that chemical events in the enzyme active site are more efficient than the corresponding processes in solution[6]. On the other hand, biomolecules are highly flexible and entropic contributions to the reaction free energy can be as important as potential energy contributions. Therefore, to model chemical events in biological systems requires both accurate potential functions and access to sufficient conformational sampling and long time-scales.

None of the existing methods alone is up to the task in general. For example, standard QM methods like Hartree-Fock (HF), Density-Functional (DFT) or Semi-Empirical (SE) Theory alone can not handle several thousands of atoms with sufficient sampling. As a consequence, many studies in the past focused only on small parts of the system, such as the active site of the protein where the reaction occurs. This however, has been shown to be insufficient due to the long range nature of the electrostatic forces and steric interactions of the active site with the environment[6–8]. The development of linear scaling methods extended the applicability of QM significantly. However, their application to large systems is still costly, not viable for many interesting systems with 10,000-100,000 atoms and not helpful when dynamical or thermodynamical properties are required, which is the case in many biological applications. Evidently, methods from different computational levels have to be combined effectively, which has been explored for the past few decades.

In the quantum chemistry community, efforts have largely been focussed on the combination of QM methods with continuum electrostatic theories, starting from Born & Onsager theories that aimed at computing the solvation free energy of charges in a polar environment. These methods have been refined over the years and can now give a reasonable description of solvation properties in an isotropic and homogeneous medium[9,10]. In this context, MM force field methods have also been combined with continuum electrostatics methods[11,12] since the number of water that has to be included in explicit solvent simulations with the periodic boundary condition often far exceeds the number of atoms in the biological molecule itself. Most of these methods are based on the Poisson-Boltzmann theory[13] and the Generalized Born model[14], although more sophisticated integral equation and density functional theories[13] have also been explored for small biomolecules.

These continuum models (CM), however, are by no means appropriate to represent the electrostatic and steric interactions of the structured environment with the active site. Therefore, Warshel and Levitt[15] proposed in 1976 to combine QM methods for the active site with MM methods for the remainder of the system. An appropriate QM-MM coupling term describes the polarization of the QM region by the charges on the MM atoms and mediate the steric interactions via covalent bonds and van der Waals contacts. Up to now, such QM/MM methods have been developed to combine many QM methods (post-HF, HF, DFT, SE) with various force fields (e.g., CHARMM, AMBER, GROMOS, ...) and have become a powerful tool for analyzing chemical events in biomolecules.

It has long been envisioned that a multiscale model can be developed for complex molecular systems in which QM/MM methods are further augmented by a continuum electrostatic model. Indeed, although efficient Ewald summation has been implemented with QM/MM potential function[16,17], the high cost and sampling challenge associated with explicit solvent simulations also becomes more transparent for QM/MM simula-

tions, especially those using high level QM methods. Practical implementations that integrate QM/MM potential with continuum electrostatics models, however, only have become available in recent years[18–20]. The major focus of this review is to summarize the key components of such QM/MM/CM models and to discuss a few relevant examples that best illustrate their value and limitations.

2 QM/MM Methods

The development of QM/MM methods in recent years has turned them into powerful predictive tool and many research groups are involved in the process; most of the recent developments have been nicely summarized in a comprehensive review[21] (see the contribution of W. Thiel). There is not one single QM/MM method, and the multitude of different implementations can be characterized by several main distinctions:

- **Additive and subtractive methods:** Subtractive models[22] apply the QM method to the active site and the MM method to the entire system, also including the active site. Since the active site region is treated by both methods, the MM contribution for the active site has to be subtracted out:

$$E = E_{MM}^{tot} + E_{QM}^{active\ site} - E_{MM}^{active\ site} \tag{1}$$

  The advantage of this method is that it allows in a simple way to also combine two different QM methods in a QM/QM' scheme or multiple methods in a QM/QM'/MM scheme, where high (e.g., DFT) and low level (SE) QM methods are combined[23,24]. The disadvantage is that the MM has to treat also the active site, which may not be straightforward when the active site has complex electronic structure (e.g., transition metal centers). The additive scheme[25], by contrast, only applies the MM to the environment of the active site, and the two regions are then coupled by a QM/MM coupling term:

$$E = E_{QM}^{active\ site} + E_{MM}^{environement} + E_{QM/MM} \tag{2}$$

  Here, no force field parameters are needed for the active site, but the description of the boundary is conceptionally more involved.

- **The treatment of the QM/MM boundary:** In many applications, this boundary dissects a covalent bond. In the simplest *link atom* approach[25], the dangling bond of the QM region is saturated by an additional hydrogen. Other approaches avoid the introduction of this artificial hydrogen. The *pseudoatom/bond* approach[26] treats the frontier functional group as a pseudo-atom with an effective one-electron potential. In most cases, a C-C single bond has to be cut and the CH_2 at the QM boundary is then substituted by a parametrized (using a pseudo-potential) pseudo-Fluorine, which models the properties of the C-C bond. The hybrid-orbital approach[27] does not substitute the boundary CH_2 group but freezes the occupation of the orbital, which represents the dangling bond. These are the most common approaches to deal with the QM/MM boundary and various variants have also been proposed[28]. Systematic studies indicate that most schemes give comparable results as far as the charges at the QM/MM boundary are carefully treated[29–31].

- **Mechanical, electrostatic and polarizable embedding:** This concerns the QM/MM coupling term and the nature of the force field. In the mechanical embedding[22,23], the MM point charges are not allowed to polarize the QM region. The interaction of the QM and MM regions is simply given by the Coulomb and van der Waals interactions[32] between the QM and MM subsystems and the interactions at the boundary, thus the QM density is *not* perturbed by the MM charges. Since biological systems are often highly charged, this method should not be used for biological applications. The electrostatic embedding[25] includes the MM charges as external point charges in the determination of the QM charge density, i.e., the QM region is polarized by the MM charges. This sounds conceptually simple, but can be an intricate matter in practice. First of all, the QM density can become too delocalized due to interactions with the point charges, which is referred to as the "spill out problem", in particular when large basis sets of plane wave bases are used. This problem can be alleviated by using a modification of the $1/r$ interaction at short distances[33]. Further, large point charges close to the QM region can overpolarize the QM density due to the artificial concentration of the MM charge at one point. Here, a charge smearing scheme can be used[28]. Finally, in the *polarizable embedding* scheme a polarizable force field instead of the fixed point charge model is used. In some cases, polarization effects from the environment can have a significant impact on the result as shown, for example, by the calculation of excitation energies in retinal proteins[34,35] (see below).

3 Sampling Reactive Events

For chemical reactions, the calculation of free energy changes and activation free energies is of ultimate interest and is still a challenge. There are several categories of techniques available.

- **Direct MD** The most straightforward way is to perform MD simulations by integrating Newton's equation of motion with either the microcanonical or canonical ensembles.[36]. The common computational technology and algorithms, however, put severe limitations in the accessible time scales. As a rule of thumb, HF and DFT methods allow to perform MD simulations in the ps regime ($\approx$ 10-50ps for 'small' QM regions of 10-50 atoms), while SE methods allow for simulation times roughly three orders of magnitude longer ($\approx$ 10-100ns for 'small' QM regions). Therefore, direct MD simulations only allow overcoming small free energy barriers of several k_BT, such as sampling of various conformers of very short peptides in water (see below). Many chemical reactions of interest have barriers on the order of 10-25 kcal/mol, and can not be meaningfully addressed with direct MD simulations, even with SE methods. Direct MD simulations, therefore, are mostly useful for equilibrating configurations of protein active sites and qualitative exploration of the structural features relevant to chemistry, such as water distributions along potential proton transfer pathways.

- **Reaction path techniques** These methods determine the Minimum Energy Path (MEP)[37] between a reactant and product state, in particular they locate the transition state (saddle point on the potential energy surface). For enthalpy driven processes, this path contains most relevant information for describing the chemical reaction of interest, in particular the relative energies of reactant, product and transition state. As

a starting point, reactant and product states have to be available. For simple reactions, an approximate MEP can be determined by the adiabatic mapping procedure[38], when a reaction coordinate is chosen and partial optimizations are carried out with the reaction coordinate set to a number of values; e.g., consider a proton transfer from an oxygen atom to a nitrogen, denote the O-H distance by d_1, the H-N distance by d_2, a reaction coordinate $d = d_1 - d_2$ can then be used to describe the reaction. For more complex reaction processes that actively involve many degrees of freedom, however, more sophisticated techniques are required. One technique available in CHARMM is called the Conjugate Peak Refinement (CPR,[39]), which starts by a straight line interpolation between reactant and product. At the line search maximum, all degrees of freedom perpendicular ('conjugate') to the initial search direction are optimized, until a minimum is found. This minimum is then connected to the reactant and product and the optimization process is iterated. A popular alternative is the Nudged elastic band method (NEB[40]), where images of the system are distributed along the search line between reactant and product and are connected by springs; the related dimer method[41] is also widely used, though more in solid state and surface physics communities.

For enthalpy driven processes, MEP based techniques can provide valuable mechanistic information. The limitations of the methods, however, are also obvious. First, the straight line interpolation does not assure to find the pathway with lowest energy [a]. Therefore, chemical intuition is necessary to include various different intermediate states, as illustrated in our study of the first proton transfer event in Bacteriorhodopsin[42] (see below). Moreover, entropic contributions are completely neglected. For example, Klähn et al.[43] showed for the reaction of a phosphate ion in the Ras-GAP complex that the total energies of reactant and product fluctuate on the order of 30 kcal/mole and the reaction barrier on the order of 6 kcal/mol, when using different protein conformations generated by classical MD simulations. In other words, the thermal motion of the protein environment makes the use of total energies in the MEP framework meaningless, which highlights the point that pursuing a high accuracy in the QM method may not be the bottleneck for meaningful QM/MM studies of many biological problems.

- **Free energy computations along reaction path** One approach for improving upon MEP results is to calculate the free energy (potential of mean force) along the MEP. For example, the MM free energy contribution along the MEP can be estimated using free energy perturbation calculations in which the QM region is frozen (or treated in a harmonic fashion) while the MM region samples the proper thermal distribution orthogonal to the MEP[44]. In the more elaborate scheme developed recently[45], the path itself can be refined based on the derivatives of the potential of mean force, which ultimately converges to a minimum free energy path. The cost of such calculations, however, can be rather high especially if high-level QM methods are used; one practical approximation is to replace the QM region by effective (or even polarizable) charges when sampling the MM degrees of freedom[46].

- **Umbrella sampling and meta-dynamics** When the reaction can be described by a

[a] Imagine connecting Munich and Milano by a rope, which will arrange along the valleys connecting Munich and Milano: however, depending on the initial placement of the rope, different pathways can be found.

number of pre-chosen "reaction coordinates", umbrella sampling techniques[47] can be used to generate the relevant potential of mean force curve/surface. The most basic technique is to add harmonic umbrella potentials at a discrete set of reaction coordinate values to overcome barriers on the potential energy surface, and various schemes have been proposed to make the process automated (adaptive) and converge quickly. For example, meta-dynamics methods[48] are adaptive umbrella sampling methods where successive Gaussians are added to avoid revisiting configurations during the sampling and therefore speeds up the convergence; the width and height of the added Gaussian functions as well as the frequency of adding the Gaussian functions can be optimized for optimal convergence[49–51]. Finally, energy can be used as a collective reaction coordinate to enhance sampling when it is difficult to determine *a priori* a set of geometrical parameters that describe the reaction[52–54].

- **Other advanced techniques** Finally, there are transition path sampling (TPS) techniques that aim to directly sample the reactive trajectory ensembles[55]. These are in principle the most rigorous framework for understanding reaction mechanisms in the condensed phase and generally do not require specifying *a priori* the reaction coordinates; it is well known that environmental degrees of freedom can be essential part of the kinetic bottleneck for many reactions in solution and biological systems. TPS has been applied in several studies of enzyme reactions[56,57], and the cost of such calculations highlights the importance of developing accurate SE methods. It should be noted that the TPS techniques in principle can also suffer from sampling issues in the path space and therefore can also benefit from using different initial guesses.

4 Semi-Empirical Methods

While the adiabatic mapping calculations can be readily applied in conjunction with HF and DFT methods, more elaborate reaction path techniques and most free energy and TPS techniques overstretch the possibilities of *ab initio* methods and are mostly applied using SE methods. The great promise of DFT methods on the one hand and the lower accuracy and limited transferability of the SE type methods, like MNDO, AM1 or PM3 on the other hand, seemed to devalue the latter type of methods; in the late 1990's they were to become obsolete in the eyes of many quantum chemist's. However, the limitations and quite involved empirical parametrization process of modern DFT methods changed also the view onto the SE methods[58]. The desire to study increasingly complex (bio)molecules and the importance of entropic contribution and sampling in studying soft matter brought a renewed interest into SE methods, especially if they can be made more robust and transferable.

Most SE methods are derived from the Hartree-Fock theory by applying various approximations resulting in, for example, the Neglect of Differential Diatomic Overlap (NDDO) type of methods; the most well-known ones being the MNDO, AM1 and PM3 models[59]. In these methods certain integrals are omitted and the remaining are treated as parameters, which are either pre-calculated from first principles or fitted to experimental data. SE methods usually have an overall accuracy lower than DFT, although this can be reversed for specific systems. In the so called specific reaction parametrization (SRP) scheme[60], a SE method (e.g., PM3) is specifically re-parametrized for the particular sys-

tem under study, which may provide a very accurate description for the reaction of interest at a level even unmatched by popular DFT methods. However, parameterization of a SRP that works well for condensed phase simulations is not as straightforward as for gas-phase applications and a large number of carefully constructed benchmark calculations are needed[61–63]. Therefore, it remains an interesting challenge to develop generally robust SE methods that properly balance computational efficiency and accuracy. Some of the more recent models include the inclusion of orthogonalization corrections in the OMx model[59], the PDDG/PM3 model[64], and the NO-MNDO model, which all generated encouraging improvements over traditional NDDO methods.

SE methods can also be derived from DFT, a development that we have focussed on over the last decade. The so called Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) method[65,66] is derived by expanding the DFT total energy functional up to second order with respect to the charge density fluctuations $\delta\rho$ around the reference density ρ_0[66] ($\rho_0' = \rho_0(\vec{r}')$, $\int' = \int d\vec{r}'$):

$$E = \sum_i^{occ} \langle \Phi_i | \hat{H}^0 | \Phi_i \rangle + \frac{1}{2} \iint' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \left. \frac{\delta^2 E_{xc}}{\delta\rho\, \delta\rho'} \right|_{\rho_0} \right) \delta\rho\, \delta\rho' - \frac{1}{2} \iint' \frac{\rho_0' \rho_0}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0]\rho_0 + E_{cc} \tag{3}$$

$\hat{H}^0 = \hat{H}[\rho_0]$ is the effective Kohn-Sham Hamiltonian evaluated at the reference density ρ_0 and the Φ_i are Kohn-Sham orbitals. E_{xc} and V_{xc} are the exchange-correlation energy and potential, respectively, and E_{cc} is the core-core repulsion energy (an extension up to third order has been presented recently[67,68]).

The (artificial) reference density ρ_0 is chosen as a superposition of densities ρ_0^α of the neutral atoms α constituting the molecular system,

$$\rho_0 = \sum_\alpha \rho_0^\alpha \tag{4}$$

and a density fluctuation $\delta\rho$, also built up from atomic contributions

$$\delta\rho = \sum_\alpha \delta\rho^\alpha, \tag{5}$$

in order to represent the ground state density

$$\rho = \rho_0 + \delta\rho. \tag{6}$$

Approximations to the three energy contributions in eq. 3 result in the final expression of the SCC-DFTB model[66]:

$$E = \sum_{i\mu\nu}^{occ} c_\mu^i c_\nu^i < \eta_\mu | \hat{H}^0 | \eta_\nu > + \frac{1}{2} \sum_{\alpha,\beta} U_{\alpha\beta}(R_{\alpha\beta}) + \frac{1}{2} \sum_{\alpha\beta} \Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta} \tag{7}$$

SCC-DFTB has been tested in detail for atomization energies, geometries and vibrational frequencies using a large set of molecules[69–71]. In terms of atomization energies, the modern NDDO type methods like PDDG/PM2 or OM2 have bee shown to be superior to SCC-DFTB, while SCC-DFTB is excellent in reproducing geometries and also predicts reasonable vibrational frequencies. It is worth emphasizing again that the SE methods are likely less accurate than modern DFT-functionals on average, although this situation can be reversed in specific cases.[b] Moreover, as discussed above, the errors introduce by neglecting the effects of dynamics and entropy can become larger than the intrinsic error of the respective electronic structure method. Nanoseconds of MD simulations are readily feasible with SE methods, while impossible with HF and DFT. Therefore, SE methods can be used in various ways to improve the quality of the computational model: (i) They can be applied as the main QM method for the initial exploration of possible reaction mechanisms after careful testing/refinement for relevant model systems; (ii) they can be used to estimate the entropic contributions of a particular mechanism while the accurate potential energy is evaluated at a higher level method[77]; (iii) they can be used as the lower level QM in either an ONIOM type multi-level[23,24] scheme or to guide the sampling in a multi-level free energy calculations.

5 The Continuum Component

While continuum approaches applied in computational materials science mostly model mechanical properties, those applied in biological simulations mainly model the dielectric response of the environment to the charge distribution of the molecule[c]. Most popular continuum electrostatics models in the biological context are based on the Poisson-Boltzmann framework.[11,13]. The Poisson equation allows to compute the electrostatic potential and electrostatic free energy for the charge distribution of the solute in the presence of a dielectric continuum (representing the dielectric response from the solvent molecules). The PB equation further includes the mobile charge distribution originating from the surrounding ions, which respond to and modulate the electrostatic potentials of the solute charges:

$$-\nabla\epsilon(x)\nabla\phi(x) - \sum_{k=1}^{N} c_k q_k e^{-q_k\phi(x)-V_k(x)} = \frac{4\pi e^2}{kT}\rho(x) \tag{8}$$

$\rho(x)$ describes the solute charge distribution, q_k the charge of the mobile ion species k, $V_k(x)$ the steric interaction of the solute with the mobile ion species k, $\epsilon(x)$ the space-dependent dielectric function and $\phi(x)$ the resulting electrostatic potential. As discussed extensively in the literature[13], the correlation between the mobile ions is ignored in the PB approach, thus PB is most reliable for monovalent ions, which fortunately fits most biological applications.

[b] Even well established methods like the hybrid DFT method B3LYP show deficiencies, which may not be widely recognized, e.g., problems with the description of extended electronic π systems[72,73], dispersion interactions[74] or electronically excited states with significant charge-separation[75,76]. These examples show that careful testing are obligatory before application to new systems, even for DFT methods.

[c] Generally, the work for cavity formation and the van der Waals interactions at the surface, the 'apolar' components of the solute-solvent interaction, need to be included as well, see[78].

In the most straightforward conceptual scheme, the QM/MM/CM treats the active site with QM, the entire biomolecule with MM and the solvent with CM. In many practical applications, however, it is sufficient to treat atoms very close to the active site (e.g., within 20 Å) with discrete MM degree of freedom that are fully flexible during the simulation; this would include explicit solvent molecules in or near the active site, which helps alleviate some of the limitations of continuum electrostatics models at the solute/solvent interface. To properly and efficiently deal with protein atoms in the continuum region, we have adapted the Generalized Solvent Boundary Condition (GSBP) scheme developed by Roux and co-workers for classical simulations[79]. Briefly, if we refer the discrete QM/MM region as the inner region while the continuum as the outer regions, the total effective potential (potential of mean force) of the system can be written as,

$$W_{GSBP} = U^{(ii)} + U^{(io)}_{int} + U^{(io)}_{LJ} + \Delta W_{np} + \Delta W^{(io)}_{elec} + \Delta W^{(ii)}_{elec}, \tag{9}$$

where $U^{(ii)}$ is the complete inner-inner potential energy, $U^{(io)}_{int}$ and $U^{(io)}_{LJ}$ are the inner-outer internal (bonds, angles, and dihedrals) and Lennard-Jones potential energies, respectively, and ΔW_{np} is the non-polar confining potential. The last two terms in Eq.9 are the core of GSBP, representing the long-range electrostatic interaction between the outer and inner regions. The contribution from distant protein charges (screened by the bulk solvent) in the outer region, $\Delta W^{(io)}_{elec}$, is represented in terms of the corresponding electrostatic potential in the inner region, $\phi^{(o)}_s(\mathbf{r}_\alpha)$,

$$\Delta W^{(io)}_{elec} = \sum_{\alpha \in inner} q_\alpha \phi^{(o)}_s(\mathbf{r}_\alpha) \tag{10}$$

The dielectric effect on the interactions among inner region atoms is represented through a reaction field term,

$$\Delta W^{(ii)}_{elec} = \frac{1}{2} \sum_{mn} Q_m M_{mn} Q_n \tag{11}$$

where $\mathbf{M}$ and Q are the generalized reaction field matrix and generalized multipole moments, respectively, in a basis set expansion.[79]

The advantage of the GSBP method lies in its ability to include these contributions explicitly while sampling configurational space of the reaction region during a simulation at minimal additional cost. The static field potential, $\phi^{(o)}_s(\mathbf{r})$, and the generalized reaction field matrix $\mathbf{M}$ are computed only once based on PB calculations and stored for subsequent simulations. The only quantities that need to be updated during the simulation are the generalized multipole moments, Q_n,

$$Q_n = \sum_{\alpha \in inner} q_\alpha b_n(\mathbf{r}_\alpha) \tag{12}$$

where $b_n(\mathbf{r}_\alpha)$ is the nth basis function at nuclear position $\mathbf{r}_\alpha$.

As described in Ref.[18], the implementation of GSBP into a combined QM/MM framework is straightforward, and involves the QM-QM and QM-MM reaction field, and the QM-static field terms. For the GSBP combined with SCC-DFTB, these terms take on a simple form because $\rho^{QM}(\mathbf{r})$ is expressed in terms of Mulliken charges.[66] Although the formulation of GSBP is self-consistent, the validity of the approach depends on many factors especially the size of the inner region and the choice of the dielectric "constant" for

the outer region. Therefore, for any specific application, the simulation protocol has to be carefully tested using relevant benchmarks such as pK_a of key residues[17,80].

An economic alternative to the GSBP approach is the charge-scaling protocol, where the partial charges for charged-residues in the MM region are scaled down based on the electrostatic potentials calculated (with PB) when the biomolecule is in vacuum vs. solution; the scaled partial charges are then used in the QM/MM simulations with the system in vacuum. In the end, PB calculations are carried out with scaled and full partial charges to complete the thermodynamic cycle. The charge-scaling approach has been successfully used in several QM/MM studies[7,81,82], although several numerical issues (e.g., treatment of residues very close to the QM region and cancellation of large contributions) render the protocol less robust than GSBP.

6 Polarizable Force Field Models

Continuum electrostatic approaches take into account a majority of the dielectric responses of the solute. The electronic polarization of the environment to changes in the QM density during chemical reaction, however, is missing when non-polarizable force fields are used as MM. This electronic polarization may give significant contributions when electrons or ions are transported over long distances and for excitation energies, where the dipole moment of the QM region changes significantly upon excitation. In the last decade, many research groups have been actively developing polarizable force fields, for which several good overviews are available (see a thematic issue that follows *J. Chem. Theory Comput.* 2007, 3, 1877).

Common approaches to describe electronic polarization effects use models based on atomic polarizabilities, a method that we have implemented to estimate the protein polarization effects on electronic excitation energies[35]. Here, the Coulomb interaction is described using atomic charges q_A and atomic polarizabilities α_A, where the induced atomic dipoles μ_A can be calculated as:

$$\mu_A = \alpha_A \left(\sum_B \mathbf{T_{AB}} q_B + \sum_C \mathbf{T_{AC}} \mu_C \right) \tag{13}$$

The first term contains the Coulomb interaction with the fixed atomic point charges, which lead to the induced dipoles. The second term describes the interaction between the induced dipoles. Note that the induced dipole moments appear on both sides of the equation. For small systems, these equations can be solved by matrix inversion techniques, for large systems they are usually solved iteratively. The tensors $\mathbf{T}$ contain a damped Coulomb interaction since for small distances the bare Coulomb $1/r$ and $1/r^3$ terms for the charge-charge and charge-dipole interactions would lead to over-polarization. Effectively, this damping is induced by smearing out the charges, i.e., by describing the atomic charge with an exponential charge distribution. A new parameter, the width 'a' of this charge distribution is therefore introduced and has to be determined during the fitting procedure.

Atomic polarizabilities can be calculated or taken from experiment; we have used values from the literature[35], where typical parameters are around 0.5 $\mathring{A}^{-3}$ for H and about 0.8-1.5 $\mathring{A}^{-3}$ for first row atoms C, N and O. However, to gain high accuracy atomic parameters have been taken to be dependent on the atomic hybridization state, e.g., the parameters for

sp^2 and sp^3 carbon differ by about 0.5 $\mathring{A}^{-3}$. This allows to account for the different polarizabilities of sp^3 carbon structures, like alkanes compared to aromatic molecules, like polyenes or benzene.

Common force charge models are parametrized in order to account for the effects of solvation implicitly. This can be done by fitting the charges to experimental data, or by calculating them using HF/6-31G*, which is known to overestimate the magnitude of charges, thereby implicitly taking the effect of solvent polarization into account. Therefore, as a first step a new charge model has to be developed in order to be consistent with an explicit treatment of polarization. We computed 'polarization free' charges by performing B3LYP/6-311G(2d,2p) calculations and fitting the charges to produce the electrostatic potential at certain points at the molecular boundary (RESP)[35]. [d] These 'polarization free charges' are computed for certain molecular fragments in gas phase, i.e. for certain chemical groups like amino acid side chains. The charges therefore already contain the mutual polarization within these fragments. Therefore, the polarization model is also restricted interactions between these fragments and not applied within one region to avoid double counting.

Critical tests include the calculation of polarizability tensors of amino acid side chains in comparison with DFT and MP2 data, and the evaluation of the polarization energy of such side chains due to a probe charge in the vicinity. The polarization model is able to reproduce the QM data with high precision[35], allowing therefore for meaningful calculations on larger systems like entire proteins.

7 Applications

In this section, we discuss three applications, to illustrate the various methodologies discussed above.

7.1 Direct QM/MM MD with periodic boundary conditions: Dynamics of peptides and proteins

The conformation of peptides and proteins depend sensitively on the proper inclusion of solvent. The conformations of small peptides in the gas phase are very different from those in solution and it is challenging to use a QM description of the peptide augmented with an implicit solvent model to model those properly. One possible approach is to include the first solvation shell explicitly[83], although finite temperature effects still need to be included, which can be problematic with a small "microsolvation" model. A physically more transparent model is to surround the peptide, treated with QM, by a box of MM water molecules and to apply periodic boundary conditions[84]. The main degrees of freedom in these peptides are the backbone torsions (ϕ, ψ) , which exhibit rotational barriers of a few kcal/mol (Fig.1). To sample the energy landscape of such systems, MD simulations in the order of 10-100 nano-seconds have to be performed, which is clearly only possible using SE methods. This also illustrates the limits of direct MD simulations, which can handle only systems with small barriers of a few kcal/mol. Linear scaling methods in combination with SE methods allow to simulate the dynamics of small proteins over several 100 ps[85].

[d] Diffuse basis functions should be avoided, since those would allow the charge density into regions far away from the molecule, which are not accessible in the condensed phase due to the environment.

However, this is still quite costly and there are not too many applications where a QM treatment of the entire protein is necessary and the dynamics on these short time-scales are the quantities of key interest.

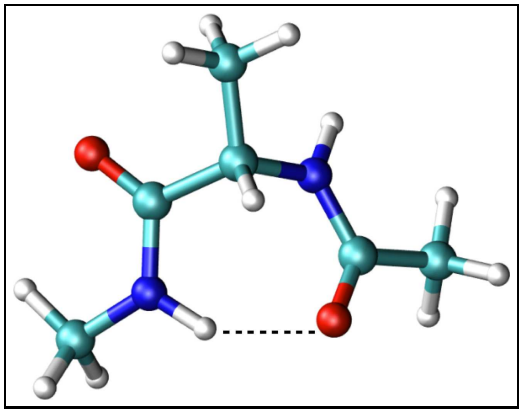


Figure 1. The lowest energy conformation C_7^{eq} of the alanine dipetide model in the gas phase. The main degrees of freedom consist of the phi and psi dihedral angles, i.e., rotations around the central C-C and C-N single bonds.

7.2 Proton transfer

Proton transfer reactions are involved in many key biological problems, most notably in acid-base catalysis and bioenergetics processes. The breaking and formation of many chemical bonds in these problems and the significant reorganization of the environment in response to the transport of charges pose great challenges to theoretical studies. Although more specialized techniques such as MS-EVB can be extremely valuable in the study of certain proton transfer problems[86], a QM/MM framework is required to introduce more flexibility in the potential energy surface, especially when the reaction involves species of complex electronic structures (e.g., transition metal ion). The diversity of proton transfer reactions also makes them ideal for illustrating the value and limitation of various QM/MM techniques.

7.2.1 Bacteriorhodopsin (bR): MEP results

For relatively localized proton transfers, for which the entropic contribution is likely small, reaction path methods can be applied. An example is the first proton transfer step in bacteriorhodopsin, where the active site involves well connected hydrogen bonding network as shown in Fig.2. It is known from experiment that entropy does not contribute to this step, therefore, we have simulated the process using SCC-DFTB QM/MM in combination with the CPR approach discussed above[42,87]. The computed barriers of 11.5-13.6 kcal/mol for different low-energy pathways are in good agreement with the experimental value of 13 kcal/mol. However, to understand this properly one has to be aware of the intrinsic error compensation in these calculations: as discussed in detail in Ref.[88], popular DFT methods

tend to underestimate proton transfer barriers by 1-4 kcal/mole. On the other hand, the inclusion of nuclear quantum effects like zero point energies would lower proton transfer barriers by roughly this amount, therefore, these two effects tend to cancel each other for a wide range of proton transfer systems.

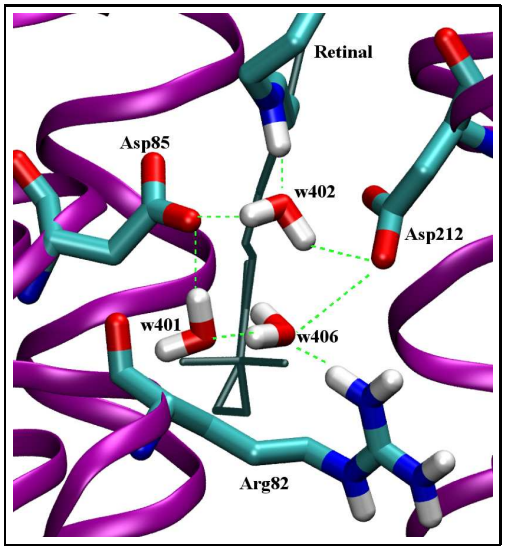


Figure 2. The active site of bacteriorhodopsin in its ground state. The first proton transfer occurs between the retinal Schiff base and the side chain Asp85.

7.2.2 Carbonic Anhydrase II : MEP vs. PMF

For many long-range proton transfers in biomolecules, however, the MEP results are likely very sensitive to the protein structure used in the calculation. More severely, the collective structural response in the protein is likely missing in the MEP calculations, which may lead to qualitatively incorrect mechanistic conclusions. A useful example in this context is the long-range proton transfer in carbonic anhydrase II (CAII), where the rate-limiting step of the catalytic cycle is a proton transfer between a zinc-bound water/hydroxide and the neutral/protonated His64 residue close to the protein/solvent interface. Since this proton transfer spans at least 8-10 Å, the transfer is believed to be mediated by the water molecules in the active site[89] (see Fig.3). Since there are multiple water wires of different length in the active site that connect the donor/acceptor groups (zinc-bound water, His 64), a question of interest is whether specific length of water wire dominates the proton transfer or all wires have comparable contributions.

First, a large number of MEPs have been collected starting from different snapshots collected from equilibrium MD simulations at the SCC-DFTB/MM level. Since essentially a positive charge is transferred over a long-distance, it was expected that the MEP energetics depend sensitively on the starting structure, which was indeed observed. For example,

when the starting structure came from a CHOH (zinc-bound water, neutral His64) trajectory, the proton transfer from the zinc-water to His64 is largely *endothermic* (on average by as much as $\sim$ 13 kcal/mol). By contrast, when the starting structure came from a COHH (zinc-bound hydroxide, protonated His64) trajectory, the same proton transfer reaction was found largely *exothermic*. As an attempt to capture the "intrinsic barrier" for the proton transfer reaction, which is known to be close to be thermoneutral experimentally,[90] we generated configurations from equilibrium MD simulations in which protons along a specific type of water wire were restrained to be equal distance from nearby heavy atoms (e.g., oxygen in water or $N\epsilon$ in His 64). In this way, the charge distribution associated with the reactive components is midway between the CHOH and COHH states, thus the active-site configuration was expected to facilitate a thermoneutral proton transfer process, which was indeed confirmed by MEP calculations using such generated configurations as the starting structure. An interesting observation is that the barriers in such "TS-reorganized" MEPs showed a steep dependence on the length of the water wire; it was small ($\sim$ 6.8$\pm$2.2 kcal/-mol) with short wires but substantially higher than the experimental value ($\sim$ 10 kcal/mol) with longer water wires (e.g., 17.4$\pm$2.0 kcal/mol for four-water wires).

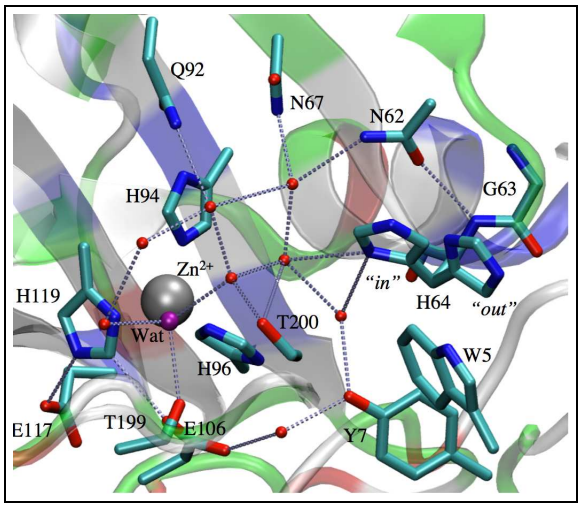


Figure 3. The active site of CAII rendered from the crystal structure (PDB ID: 2CBA[89]). All dotted lines correspond to hydrogen-bonding interactions with distances $\leq$3.5 Å. The proton acceptor, His64, is resolved to partially occupy both the "in" and "out" rotameric states.

This steep wire-length dependence is in striking contrast with the more rigorous PMF calculations.[91,92] In the PMF calculations, a collective coordinate[93] was used to monitor the progress of the proton transfer without enforcing specific sequence of events involving individual protons along the wire; the use of a collective coordinate is important because this allows averaging over different water wire configurations, which is proper since the life-time of various water wires is on the pico-second time scale,[18,20] much faster than the time scale of the proton transfer (μs).[90] In the PMF calculations, the wire-length dependence was examined by comparing results with different His 64 orientations ("in" and

"out", which is about 8 and 11 Å from the zinc, respectively); both configurations were found to involve multiple lengths of water wires but different relative populations. The two sets of PMF calculations produced barriers of very similar values, which suggested that the length of the water wire (or orientation of the acceptor group) is unlikely essential to the proton transfer rate. Further analysis of the configurations sampled in the MEP simulations suggested that the MEP results artificially favored the concerted proton transfers, which correlate to significant distance dependence. As discussed above, to generate the "TS-reorganized" configurations, all transferring protons along the wire were constrained to be half-way between the neighboring heavy atoms; therefore, such sampled protein/-solvent configurations would favor a concerted over step-wise proton transfers. Although *all* atoms in the inner region are allowed to move in the MEP searches, the local nature of MEPs does not allow collective reorganization of the active site residues/solvent molecules thus the "memory" of the sampling procedure is not erased.

Therefore, the CAII example clearly illustrates that care must be exercised when using MEP to probe the mechanism of chemical reactions in biomolecules, especially when collective rearrangements in the environment are expected (e.g., reactions involving charge transport). Along the same line, the GSBP based QM/MM/CM framework was found to be particularly attractive in the CAII studies for maintaining the proper solvent configurations and sidechain orientations in the active site, as compared to Ewald based SCC-DFTB/MM simulations[18,80], at a faction of the computational cost. Ignoring the bulk solvation effect, for example, was found to lead to unphysical orientations of the functionally important His64 residue.

7.2.3 Membrane proteins

A particularly exciting area for which the multiscale QM/MM/CM approach is suited concerns proton translocation across membrane proteins, where a proper and efficient treatment of the heterogeneous protein/solvent/membrane environment is particularly important, such as in bacteriorhodopsin and cytochrome c oxidase. The GSBP framework also allows one to incorporate the effect of membrane potential[94], which plays a major role in bioenergetics, in a numerically efficient manner. Using the SCC-DFTB/MM/GSBP protocol with a relatively small inner region ($\sim$ 30Å$\times$30Å$\times$50Å) and dielectric membrane model[93], we were able to reproduce the water wire configurations in the interior of aquaporin in good agreement with the much more elaborate MD simulations using four copies of aquaporin embedded in an explicit lipid bilayer. Ignoring the GSBP contributions, however, led to very different water distributions, which highlights the importance and reliability of the multiscale framework. In a recent study[95], the same framework was also found semi-quantitatively successful in predicting pK_a of titritable groups in the interior of bacteriorhodopsin and cytochrome c oxidase, which are extremely challenging and relevant benchmark for studying proton transfer systems in general[96]. Finally, the SCC-DFTB/MM/GSBP studies of the proton release group (PRG) in bacteriorhodopsin[97] led to the key insight that the PRG is not a protonated water cluster as proposed in a series of recent IR studies[98,99]; rather, the PRG is a pair of conserved Glutamate bonded together with a delocalized proton (see Fig.4), and it is the delocalization of this "intermolecular proton bond" that leads to the unusual IR signature found in experiments[98,99].

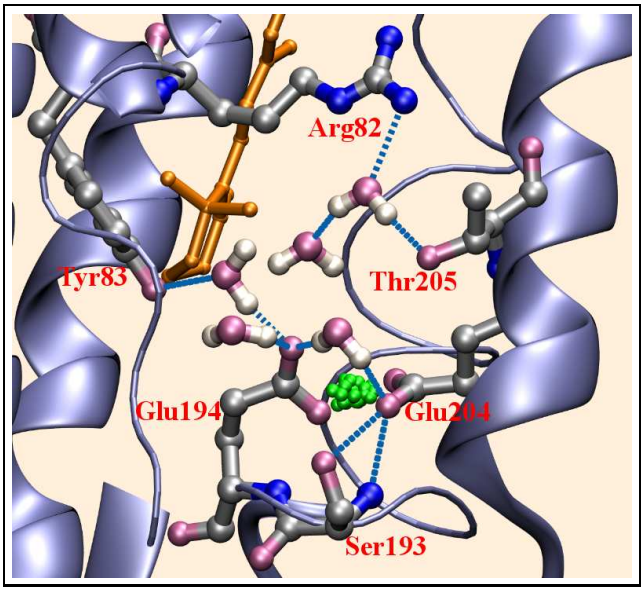


Figure 4. SCC-DFTB/MM-GSBP simulations indicate that the stored proton in the proton pump bacteriorhodopsin is delocalized (green spheres) between a pair of conserved glutamate residues rather than among the active site water molecules.

7.3 Excited states properties

The accurate determination of excited states properties is a challenging task for quantum chemical methods in general. This holds true in particular for the chromophore in retinal proteins (like bR), a polyene chain linked via a Schiff-base (NH) group to the protein backbone[76, 73, 100] (see Fig.5). Due to its extended and highly-correlatedπ-electron system, retinal is highly polarizable and undergoes a large change in dipole moment upon excitation, therefore, protein polarization effects may become important for an accurate description of excited state properties.

Standard QM/MM calculations using only an electrostatic embedding scheme do not take the (electronic) polarization response of the protein environment into account, which is different for ground and excited states due to the change of the dipole moment upon excitation. [e] In the case of retinal, the dipole in the excited state is about 10 Debye larger than in the ground state, therefore, MM polarization stabilizes the excited state more than the ground state, leading to an effective red-shift in excitation energies.

Indeed, QM/MM electrostatic embedding calculations tend to overestimate the excitation energy. While the experimental absorption maximum is at 2.18 eV, MRCI QM/MM calculations estimate it to be 2.34 eV, other methods predict even more blue shifted values[73]. There are many factors that contribute to the computational uncertainty, one of which being the intrinsic accuracy of the applied QM method. Other factors are related to the QM/MM coupling and the electrostatic treatment of the environment. For example,

[e]They of course can take the 'ionic' response into account, i.e., the relaxation of the protein structure, which also leads to a change in the electrostatic field from the MM environment.

different force field models (like AMBER and CHARMM) use different point charge models, which can lead to differences in the excitation energies on the order of 0.05 eV[35]. In many applications, only the protein is included in the MM treatment, the larger environment including membrane and bulk water is neglected. This effect can be estimated with a linearized version of the Poisson Boltzmann equation 8 in the charge scaling[81] approach as discussed above. Estimating excitation energies with and without charges scaling results again in differences of about 0.05 eV.

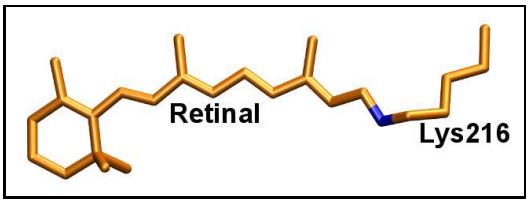


Figure 5. The retinal chromophore in the all-$trans$ conformation, as in the bR ground state. The blue color indicates the Schiff base (NH) group, from which the proton is transferred in the first step to Asp85.

Using a polarizable model, the ground state charge distribution in the MM region is determined using eq. 13. The resulting charges may be different from those in the regular force field models, because they are computed in response to the actual electrostatic field of the protein with retinal in the ground state. This charge distribution leads to vertical excitation energies about 0.07 eV red-shifted compared to those from the CHARMM force field[35]. In the same way, a different set of MM charges can be determined for the case where retinal is in its excited state. This change in the electrostatic environment leads to a further red shit of 0.07 eV, which is due to the different MM polarization in the ground and excited states. The total red-shift with respect to the CHARMM charges is 0.14eV, showing that protein polarization can have a significant impact on excitation energies in those cases, where the dipole moment of the chromophore changes significantly upon excitation.

A different approach to estimate the effect of polarization is to use a low level QM method instead of the polarizable MM region. We have used such a QM/QM'/MM approach, applying charge scaling, a MRCI method for the QM region containing the retinal chromophore and a DFT methods for 300 atoms around the chromophore in the QM' region to benchmark the polarizable MM model[34]. This study showed that the well-calibrated polarizable MM model gives nearly the same results as the QM' region. However, the 300 atom QM' region leads only to roughly 50% of the red-shift, showing that a large MM region contributes to the polarization effect.

8 Summary

In the last decade, many variants of multiscale methods have been developed to study chemical events in complex environments in materials science, chemistry and biology. The specific design of such methods depends very much on the properties of the investigated system and the problem in hand. Biological systems are characterized by their high degree of structural flexibility and the long-range nature of the electrostatic forces, which

are essential to the understanding of biological functions. Therefore, the main emphasis in methods development in the biological context lies in the accurate representation of electrostatics and algorithms to tackle the sampling problem. In this article, we have discussed QM/MM algorithms embedded into an implicit electrostatic environment, which is modeled based on the Poisson-Boltzmann equation. For many applications, the representation of the MM environment by fixed point charges may be appropriate, however, in cases where the electrostatic properties in the QM region change significantly, a polarizable MM representation is likely required. Thermal fluctuations, on the other hand, can lead to a significant contribution to the free energies that characterize the chemical reaction. Accordingly, expensive QM methods often have to be substituted by more efficient, although less accurate ones. We have described applications using various approximations for the QM region. For the determination of excitation energies, high level QM methods have to be applied, while for the study of proton transfer events, DFT and approximate SCC-DFTB can lead to a balanced treatment allowing to draw meaningful conclusions about the reaction mechanism and energetics. In some cases, the neglect of thermal fluctuations would even lead to much larger errors than the use of lower accuracy QM methods. Therefore, studying biological systems requires applying a multitude of methods and calculating multiple experimental observables to reach reliable mechanistic conclusions.

Acknowledgments

We are indebted to our collaborators for their contributions, without which the work described here can't be accomplished. Supports from the National Science Foundation, National Institutes of Health, Alfred P. Sloan Foundation, DFG and computational resources from the National Center for Supercomputing Applications at the University of Illinois are greatly appreciated.

References

1. M. Karplus and J. A. McCammon, *Molecular dynamics simulations of biomolecules*, Nat. Struct. Mol. Biol., **9**, 646–652, 2002.
2. W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glattli, P. H. Hünenberger, M. A. Kastenholz, C. Ostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, and H. B. Yu, *Biomolecular modeling: Goals, problems, perspectives*, Angew. Chem. Int. Ed., **45**, 4064–4092, 2006.
3. A. D. MacKerell Jr., D. Bashford, M. Bellot, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III., B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins*, J. Phys. Chem. B, **102**, 3586–3616, 1998.
4. W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, *Biomolecular Simulation: The GROMOS Manual and User Guide.*, vdf Hochschulverlag, ETH Zürich, Switzerland, 1996.

5. J. W. Ponder and D. A. Case, *Force fields for protein simulations*, Adv. Prot. Chem., **66**, 27, 2003.
6. A. Warshel, *Computer simulations of enzyme catalysis: Methods, Progress and Insights*, Annu. Rev. Biophys. Biomol. Struct., **32**, 425–443, 2003.
7. Qiang Cui and Martin Karplus, *Catalysis and specificity in enzymes: A study of triosephosphate isomerase (TIM) and comparison with methylglyoxal synthase (MGS)*, Adv. Prot. Chem., **66**, 315–372, 2003.
8. J. L. Gao, S. H. Ma, D. T. Major, K. Nam, J. Z. Pu, and D. G. Truhlar, *Mechanisms and free energies of enzymatic reactions*, Chem. Rev., **106**, 3188–3209, 2006.
9. M. Cossi, V. Barone, R. Cammi, and J. Tomasi, *Ab initio study of solvated molecules: A new implementation of the polarizable continuum model*, Chem. Phys. Lett., **255**, 327–335, 1996.
10. C. J. Cramer and D. G. Truhlar, *Implicit solvation models: Equilibria, structure, spectra, and dynamics*, Chem. Rev., **99**, 2161–2200, 1999.
11. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, *Electrostatics of nanosystems: Application to microtubules and the ribosome*, Proc. Acad. Natl. Sci. USA, **98**, 10037–10041, 2001.
12. M. Feig and C. L. Brooks, *Recent advances in the development and application of implicit solvent models in biomolecule simulations*, Curr. Opin. Struct. Biol., **14**, 217–224, 2004.
13. J. P. Hansen and I. R. McDonald, *Theory of simple liquids, 3rd Ed.*, Academic Press, London, UK, 2006.
14. D. Bashford and D. A. Case, *Generalized born models of macromolecular solvation effects*, Annu. Rev. Phys. Chem., **51**, 129–152, 2000.
15. Warshel, A. and Levitt, M., *Theoretical Studies of Enzymic Reactions*, J. Mol. Biol., 1976.
16. K. Nam, J. Gao, and D. M. York, *An efficient linear-scaling ewald method for long-range electrostatic interactions in combined QM/MM calculations*, J. Chem. Theo. Comp., **1**, 2–13, 2005.
17. D. Riccardi, P. Schaefer, and Q. Cui, *pK_a calculations in solution and proteins with QM/MM free energy perturbation simulations: A quantitative test of QM/MM protocols*, J. Phys. Chem. B, **109**, 17715–17733, 2005.
18. P. Schaefer, D. Riccardi, and Q. Cui, *Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules*, J. Chem. Phys., **123**, 014905, 2005.
19. B. A. Gregersen and D. M. York, *Variational Electrostatic projection (VEP) methods for efficient modeling of the macromolecular electrostatic and solvation environment in activated dynamics simulations*, J. Phys. Chem. B, **109**, 536–556, 2005.
20. D. Riccardi, P. Schaefer, Y. Yang, H. Yu, N. Ghosh, X. Prat-Resina, Peter Konig, G. Li, D. Xu, H. Guo, M. Elstner, and Qiang Cui, *Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes (Feature Article)*, J. Phys. Chem. B, **110**, 6458–6469, 2006.
21. H. M. Senn and W. Thiel, *QM/MM studies of enzymes*, Curr. Opin. Chem. Biol., **11**, 182–187, 2007.
22. F. Maseras and K. Morokuma, *IMOMM - A new integrated ab initio plus molecular mechanics geometry optimization scheme of equilibrium structures and transition states*, J. Comp. Chem., **16**, 1170–1179, 1995.

23. M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, *ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition*, J. Phys. Chem., **100**, 19357–19363, 1996.
24. Q. Cui, H. Guo, and M. Karplus, *Combining ab initio and density functional theories with semiempirical methods*, J. Chem. Phys., **117**, 5617–5631, 2002.
25. M. J. Field, P. A. Bash, and M. Karplus, *A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations*, J. Comput. Chem., **11**, 700–733, 1990.
26. Y. Zhang, T.S. Lee, and W. Yang, *A pseudobond approach to combining quantum mechanical and molecular mechanical methods*, J. Chem. Phys., **110**, 46–54, 1999.
27. J. Gao, P. Amara, C. Alhambra, and M. J. Field, *A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations*, J. Phys. Chem. A, **102**, 4714–4721, 1998.
28. D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chattfield, M. Hodošček, and B. R. Brooks, *Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method*, J. Chem. Phys., **117**, 10534–10547, 2002.
29. N. Reuter, A. Dejaegere, B. Maigret, and M. Karplus, *Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches*, J. Phys. Chem. A, **104**, 1720–1735, 2000.
30. I. Antes and W. Thiel, *Adjusted Connection Atoms for Combined Quantum Mechanical and Molecular Mechanical Methods.*, J. Phys. Chem. A, **103**, 9290, 1999.
31. P. H. König, M. Hoffmann, Th. Frauenheim, and Q. Cui, *A critical evaluation of different QM/MM frontier treatments using SCC-DFTB as the QM method*, J. Phys. Chem. B, **109**, 9082–9095, 2005.
32. D. Riccardi, G. Li, and Q. Cui, *The importance of van der Waals interactions in QM/MM simulations*, J. Phys. Chem. B, **108**, 6467–6478, 2004.
33. A. Laio, J. VanderVondele, and U. Rothlisberger, *A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations*, j. Chem. Phys., **116**, 6941–6947, 2002.
34. M. Wanko, M. Hoffmann, T. Frauenheim, and M. Elstner, *Effect of polarization on the opsin shift in rhodopsins. 1. A combined QM/QM/MM model for bacteriorhodopsin and pharaonis sensory rhodopsin II*, J. Phys. Chem. B, **112**, 11462–11467, 2008.
35. M. Wanko, M. Hoffmann, J. Frahmcke, T. Frauenheim, and M. Elstner, *Effect of polarization on the opsin shift in rhodopsins. 2. empirical polarization models for proteins*, J. Phys. Chem. B, **112**, 11468–11478, 2008.
36. Daan Frenkel and Berend Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, San Diego, London, 2002.
37. D. J. Wales, *Energy Landscapes*, Cambridge University Press, 2003.
38. T. Siomonson, *Computational biochemistry and biophysics*, Marcel Dekker, Inc., 2001.
39. S. Fischer and M. Karplus, *Conjugate Peak Refinement : an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom.*, Chem. Phys. Lett., **194**, 252–261, 1992.
40. G. Henkelman, B. P. Uberuaga, and H. Jónsson, *Climbing image nudged elastic band*

method for finding saddle points and minimum energy paths, J. Chem. Phys., **113**, 9901–9904, 2000.
41. G. Henkelman and H. Jónsson, *A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives*, J. Chem. Phys., **111**, 7010–7022, 1999.
42. A. Bondar, S. Fischer, J. C. Smith, M. Elstner, and S. Suhai, *Key role of electrostatic interactions in bacteriorhodopsin proton transfer*, Journal of the American Chemical Society, **126**, 14668–14677, 2004.
43. M. Klahn, S. Braun-Sand, E. Rosta, and A. Warshel, *On possible pitfalls in ab initio quantum mechanics/molecular mechanics minimization approaches for studies of enzymatic reactions*, J. Phys. Chem B, **109**, 15645, 2005.
44. Y. K. Zhang, H. Y. Liu, and W. T. Yang, *Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface*, J. Chem. Phys., **112**, 3483–3492, 2000.
45. H. Hu, Z. Y. Lu, and W. T. Yang, *QM/MM minimum free-energy path: Methodology and application to triosephosphate isomerase*, J. Chem. Theo. Comp., **3**, 390–406, 2007.
46. H. Hu and W. T. Yang, *Free Energies of Chemical Reactions in Solution and in Enzymes with Ab Initio Quantum Mechanics/Molecular Mechanics Methods*, Annu. Rev. Phys. Chem., **59**, 573–601, 2008.
47. G.M. Torrie and J.P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*, J. Comp. Phys., **23**, 187–199, 1977.
48. A. Laio and M. Parrinello, *Escaping free energy minima*, Proc. Nat. Acad. Sci. USA, **99**, 12562–12566, 2002.
49. A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello, *Assessing the accuracy of metadynamics*, J. Phys. Chem. B, **109**, 6714–6721, 2005.
50. A. Barducci, G. Bussi, and M. Parrinello, *Well-tempered metadynamics: A smoothly converging and tunable free-energy method*, Phys. Rev. Lett., **100**, 020603, 2008.
51. D. H. Min, Y. S. Liu, I. Carbone, and W. Yang, *On the convergence improvement in the metadynamics simulations: A Wang-Landau recursion approach*, J. Chem. Phys., **126**, 194104, 2007.
52. H. Li, D. Min, Y. Liu, and W. Yang, *Essential energy space random walk via energy space metadynamics method to accelerate molecular dynamics simulations*, J. Chem. Phys., **127**, 094101, 2007.
53. Y. Q. Gao, *An integrate-over-temperature approach for enhanced sampling*, J. Chem. Phys., **128**, 064105, 2008.
54. D. Hamelberg, J. Mongan, and J. A. Mccammon, *Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules*, J. Chem. Phys., **120**, 11919–11929, 2004.
55. P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Transition path sampling: Throwing ropes over rough mountain passes, in the dark*, Annu. Rev. Phys. Chem., **53**, 291–318, 2002.
56. R. Crehuet and M. J. Field, *A transition path sampling study of the reaction catalyzed by the enzyme chorismate mutase*, J. Phys. Chem. B, **111**, 5708–5718, 2007.
57. S. Saen-oon, S. Quaytman-Machleder, V. L. Schramm, and S. D. Schwartz, *Atomic*

detail of chemical transformation at the transition state of an enzymatic reaction, Proc. Natl. Acad. Sci. USA, **105**, 16543–16548, 2008.
58. M. Elstner, T. Frauenheim, J. McKelvey, and G. Seifert, *Density functional tight binding: Contributions from the American chemical society symposium*, J. Phys. Chem. A, **111**, 5607–5608, 2007.
59. W. Thiel, *Perspectives on semiempirical molecular orbital theory*, Adv. Chem. Phys., **93**, 703–757, 1996.
60. I. Rossi and D. G. Truhlar, *Parameterization of NDDO wavefunctions using genetic algorithm*, Chem. Phys. Lett., **233**, 231–236, 1995.
61. Q. Cui and M. Karplus, *QM/MM Studies of the Triosephosphate Isomerase (TIM) Catalyzed Reactions: Verification of Methodology and Analysis of the Reaction Mechanisms*, J. Phys. Chem. B, **106**, 1768–1798, 2002.
62. K. Nam, Q. Cui, J. Gao, and D. M. York, *A specific reaction parameterization for the AM1/d Hamiltonian for transphosphorylation reactions*, J. Chem. Theo. Comp., **3**, 486–504, 2007.
63. Yang Yang, Haibo Yu, Darrin M. York, Marcus Elstner, and Qiang Cui, *Description of phosphate hydrolysis reactions with the Self-Consistent-Charge Tight-Binding-Density-Functional (SCC-DFTB) theory 1. Parameterization*, J. Chem. Theo. Comp., **In press**, 2008.
64. M. P. Repasky, J. Chandrasekhar, and W. L. Jorgensen, *PDDG/PM3 and PDDG/MNDO: Improved semiempirical methods*, J. Comp. Chem., **23**, 1601–1622, 2002.
65. Porezag, D., Frauenheim, T., Köhler, T., Seifert, G., and Kaschner, R., *construction of tight-binding-like potentials on the basis of density functional theory - application to carbon*, Phys. Rev. B, **51**, 12947–12957, 1995.
66. M. Elstner, D. Porezag, G. Jungnickel, J. Elstner, M. Haugk, Th. Frauenheim, S. Suhai, and G. Seifert, *Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties*, Phys. Rev. B, **58**, 7260–7268, 1998.
67. M. Elstner, *SCC-DFTB: What is the proper degree of self-consistency*, Journal of Physical Chemistry A, **111**, 5614–5621, 2007.
68. Yang, Y., Yu, H., York, D., Cui, Q., and Elstner, M., *Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method: Third-Order Expansion of the Density Functional Theory Total Energy and Introduction of a Modified Effective Coulomb Interaction*, J. Phys. Chem. A, **111**, 10861–10873, 2007.
69. T. Kruger, M. Elstner, P. Schiffels, and Th. Frauenheim, *Validation of the density functional based tight-binding approximation method for the calculation of reaction energies and other data*, J. Chem. Phys., **122**, 114110, 2005.
70. K. W. Sattelmeyer, J. Tirado-Rives, and W. L. Jorgensen, *Comparison of SCC-DFTB and NDDO-based semiempirical molecular orbital methods for organic molecules*, Journal of Physical Chemistry A, **110**, 13551–13559, 2006.
71. N. Otte, M. Scholten, and W. Thiel, *Looking at self-consistent-charge density functional tight binding from a semiempirical perspective*, Journal of Physical Chemistry A, **111**, 5751–5755, 2007.
72. A. Bondar, S. Suhai, S. Fischer, J. C. Smith, and M. Elstner, *Suppression of the back proton-transfer from Asp85 to the retinal Schiff base in bacteriorhodopsin: A theoretical analysis of structural elements*, Journal of Structural Biology, **157**, 454–469,

2007.
73. M. Wano, M. Hoffmann, P. Strodel, A. Koslowski, W. Thiel, F. Neese, T. Frauenheim, and M. Elstner, *Calculating absorption shifts for retinal proteins: Computational challenges*, J. Phys. Chem. B, **109**, 3606–3615, 2005.
74. M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, and E. Kaxiras, *Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density functional-theory based treatment*, Journal of Chemical Physics, **114**, 5149–5155, 2001.
75. A. Dreuw, J. L. Weisman, and M. Head-Gordon, *Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange*, Journal of Chemical Physics, **119**, 2943–2946, 2003.
76. M. Wanko, M. Garavelli, F. Bernardi, T. A. Niehaus, T. Frauenheim, and M. Elstner, *A global investigation of excited state surfaces within time-dependent density-functional response theory*, Journal of Chemical Physics, **120**, 1674–1692, 2004.
77. F. Claeyssens, J. N. Harvey, F. R. Manby, R. A. Mata, A. J. Mulholland, K. E. Ranaghan, M. Schütz, S. Thiel, W. Thiel, and H. J. Werner, *High-accuracy computation of reaction barriers in enzymes*, Angew. Chim. Intl. Ed., **45**, 6856–6859, 2006.
78. J. A. Wagoner and N. A. Baker, *Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms*, Proc. Nat. Acad. Sci. USA, **103**, 8331–8336, 2006.
79. W. Im, S. Berneche, and B. Roux, *Generalized solvent boundary potential for computer simulations*, J. Chem. Phys., **114**, 2924–2937, 2001.
80. D. Riccardi and Q. Cui, *pK_a analysis for the zinc-bound water in Human Carbonic Anhydrase II: benchmark for "multi-scale" QM/MM simulations and mechanistic implications*, J. Phys. Chem. A, **111**, 5703–5711, 2007.
81. A. R. Dinner, X. Lopez, and M. Karplus, *A charge-scaling method to treat solvent in QM/MM simulations*, Theoretical Chemistry Accounts, **109**, 118–124, 2003.
82. G. Li, X. Zhang, and Q. Cui, *Free Energy Perturbation Calculations with Combined QM/MM Potentials Complications, Simplifications, and Applications to Redox Potential Calculations*, J. Phys. Chem. B, **107**, 8643–8653, 2003.
83. W. G. Han, K. J. Jalkanen, M. Elstner, and S. Suhai, *Theoretical study of aqueous N-acetyl-L-alanine N'-methylamide: Structures and Raman, VCD, and ROA spectra*, Journal of Physical Chemistry B, **102**, 2587–2602, 1998.
84. H. Hu, M. Elstner, and J. Hermans, *Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution*, Proteins: Structure, Function and Genetics, **50**, 451–463, 2003.
85. H. Liu, M. Elstner, E. Kaxiras, T. Frauenheim, J. Hermans, and W. Yang, *Quantum mechanics simulation of protein dynamics on long timescale*, Proteins: Structure, Function and Genetics, **44**, 484–489, 2001.
86. J. M. J. Swanson, C. M. Maupin, H. Chen, M. K. Petersen, J. Xu, Y. Wu, and G. A. Voth, *Proton solvation and transport in aqueous and biomolecular systems: insights from computer simulations*, J. Phys. Chem. B, **111**, 4300–4314, 2007.
87. A. Bondar, M. Elstner, S. Suhai, J. C. Smith, and S. Fischer, *Mechanism of primary proton transfer in bacteriorhodopsin*, Structure, **12**, 1281–1288, 2004.

88. M. Elstner, *The SCC-DFTB method and its application to biological systems*, Theor. Chem. Acc., **116**, 316–325, 2006.
89. K. Håkansson, M. Carlsson, L. A. Svensson, and A. Liljas, *Structure of native and apo carbonic anhydrase II and structure of some its anion-ligand complexes*, J. Mol. Biol., **227**, 1192–1204, 1992.
90. D. N. Silverman, *Proton transfer in carbonic anhydrase measured by equilibrium isotope exchange*, Methods in Enzymology, **249**, 479–503, 1995.
91. D. Riccardi, P. König, X. Prat-Resina, H. Yu, M. Elstner, T. Frauenheim, and Q. Cui, *"Proton holes" in long-range proton transfer reactions in solution and enzymes: A theoretical analysis*, J. Am. Chem. Soc., **128**, 16302–16311, 2006.
92. D. Riccardi, P. Koenig, H. Guo, and Q. Cui, *Proton Transfer in Carbonic Anhydrase Is Controlled by Electrostatics Rather than the Orientation of the Acceptor*, Biochem., **47**, 2369–2378, 2008.
93. P. H. Konig, N. Ghosh, M. Hoffmann, M. Elstner, E. Tajkhorshid, Th Frauenheim, and Q. Cui, *Toward theoretical analyis of long-range proton transfer kinetics in biomolecular pumps*, Journal of Physical Chemistry A, **110**, 548–563, 2006.
94. B. Roux, *Influence of the membrane potential on the free energy of an intrinsic protein*, Biophys. J., **73**, 2980–2989, 1997.
95. N. Ghosh, X. Prat-Resina, M. Gunner, and Q. Cui, *Microscopic pK_a analysis of Glu 286 in Cytochrome c Oxidase (Rhodobacter sphaeroids): towards a calibrated molecular model*, Biochem., **Submitted**, 2008.
96. M. Kato, A. V. Pisliakov, and A. Warshel, *The barrier for proton transport in Aquaporins as a challenge for electrostatic models: The role of protein relaxation in mutational calculations*, Proteins: Struct. Funct. Bioinfor., **64**, 829–844, 2006.
97. P. Phatak, N. Ghosh, H. Yu, M. Elstner, and Q. Cui, *Amino acids with an intermolecular proton bond as proton storage site in bacteriorhodopsin*, Proc. Acad. Natl. Sci. U.S.A., **In press**, 2008.
98. F. Garczarek, L. S. Brown, J. K. Lanyi, and K. Gerwert, *Proton binding within a membrane protein by a protonated water cluster*, Proc. Acad. Natl. Sci. U.S.A., **102**, 3633–3638, 2005.
99. F. Garczarek and K. Gerwert, *Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy*, Nature, **439**, 109–112, 2006.
100. M. Hoffmann, M. Wanko, P. Strodel, P. H. Konig, T. Frauenheim, K. Schulten, W. Thiel, E. Tajkhorshid, and M. Elstner, *Color tuning in rhodopsins: The mechanism for the spectral shift between bacteriorhodopsin and sensory rhodopsin II*, J. Am. Chem. Soc., **128**, 10808–10818, 2006.

Application of Residue-Based and Shape-Based Coarse Graining to Biomolecular Simulations*

Peter L. Freddolino[1,2,4], Anton Arkhipov[1,3,4], Amy Y. Shih[2,4], Ying Yin[3,4], Zhongzhou Chen[3,4], and Klaus Schulten[2,3,4]

[1] The authors contributed equally

[2] Center for Biophysics and Computational Biology

[3] Department of Physics

[4] Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801

Because many systems of biological interest operate on time and/or length scales that are inaccessible to atomistic molecular dynamics simulations, simplified representations of biomolecular systems are often simulated, a practice known as coarse-graining. We review two modern techniques for coarse-graining biomolecular systems, and describe several example systems where each of these techniques has been successfully applied. Promising avenues for future work on refining coarse-graining methods are also discussed.

1 Introduction

A vast array of problems currently addressed by computer simulations, including biological systems, involve the analysis of properties on long time and length scales derived from simulations on relatively short time and length scales[1]. Although these techniques can provide a great deal of insight on the processes under study, traditional simulations of this type are limited in scope by their computational costs, which impose an upper limit on the time scale that can be studied (currently in the nanosecond range, for biological systems[2]). This limitation has lead to the development of a wide variety of techniques attempting to provide longer time and length scale information than traditional (usually atomistic) simulations, many of which fall into the category of coarse graining. In the broadest possible sense, the term "coarse graining" (CG) can be used to refer to any simulation technique in which a simulated system is simplified by clustering several subcomponents of it into one component, thus effectively reducing the computational complexity by removing both degrees of freedom and interactions from the system. The fundamental assumption behind such techniques is that by eliminating insignificant degrees of freedom, one will be able to obtain physically correct data on the properties of a system over longer time scales than would otherwise be achievable[3].

A wide variety of coarse graining methods for biological systems currently exist, ranging in some sense from united-atom models to elastic network models. We focus on the principles and applications of two classes of biological coarse graining, namely residue-based and shape-based coarse graining. Residue-based CG is a broad family of methods

*Reprinted from Peter L. Freddolino, Anton Arkhipov, Amy Y. Shih, Ying Yin, Zhongzhou Chen, and Klaus Schulten. Application of residue-based and shape-based coarse graining to biomolecular simulations. In Gregory A. Voth, editor, Coarse-Graining of Condensed Phase and Biomolecular Systems, chapter 20, pp. 299-315. Chapman and Hall/CRC Press, Taylor and Francis Group, 2008.

in which clusters of 10-20 covalently bonded atoms are represented by one bead; it is a fairly natural and common method for coarse graining when a speedup of 1-2 orders of magnitude over all-atom simulations is required. Shape-based CG is a method recently developed in our group which uses a neural network algorithm to assign CG beads to domains of a protein, efficiently reproducing the shape of the protein with a minimal number of particles. Interactions between beads are then parameterized from all-atom simulations of the bead components. In this chapter we present a summary of both methods, along with exemplary applications of residue-based CG to two lipid-protein systems involving large-scale conformational changes, and of shape-based CG to the mechanical properties of polymeric systems.

2 Residue-Based Coarse Graining

The most natural (and frequently used) method for coarse-graining a biological system is to assign sections of each biological molecule (or monomer, in the case of a biopolymer) with similar chemical properties and spatial location to a "bead", and then treat the coarse grained system as an ensemble of beads. This type of description is henceforth referred to as "residue-based coarse graining". For example, in one possible description of a protein each amino acid residue would be represented by two beads, one representing the backbone atoms and a second (different for each residue type) representing the side chain atoms[4,5]. An example of residue-based coarse graining is shown in Figure 1.

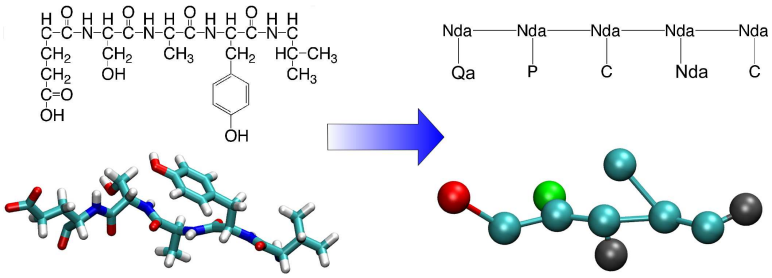


Figure 1. Structure of the polypeptide ESAYV in all-atom (left) and residue-based CG (right) representations.

While in principle similar to the united-atom models common in the early stages of molecular dynamics[6], modern residue-based CG methods are generally geared toward much longer timescales, and are thus coarser. The strategy of making a cluster of connected heavy atoms the unit particle, rather than atoms or heavy atoms, permits a longer timestep and thereby yields a larger reduction in computational effort than united atom models, but obviously carries a commensurate loss of detail. Recent interest in residue-based coarse graining has emerged in the field of lipid simulations, where several groups have developed CG lipid models either by attempting to reconstruct the forces observed in all-atom MD[7–11] or by using a created potential with parameters tuned to match experimental thermodynamic data[12–14]. In both of these cases, the coarse graining process maps approximately 10 atoms to one coarse grained particle ("bead"), and the resultant

CG model reproduced both the physical properties and (to the extent that they are experimentally known) assembly mechanisms of bilayers, micelles, and other lipid aggregates on microsecond timescales. Similar efforts have recently been extended to proteins, including simulation of protein-lipid assemblies[4,15] and protein folding[16].

2.1 Interaction potentials for residue-based CG

In the broadest sense, the forcefields used in residue-based CG models tend to fall into one of two categories, either being derived phenomenologically or through MD-based parameterization. The former approach, exemplified by the lipid-water forcefields of Marrink and co-workers[12–14] and by the more recent MARTINI forcefield[17], involves partitioning clusters of atoms into abstract "types" based on their physical properties (for example, polarity and ability to hydrogen bond); the interactions between beads are then parameterized to reproduce experimental data such as partition energies[13]. The latter approach is a direct analogue of parameterization of all-atom MD models from quantum mechanical calculations; here, all-atom simulations are performed on some system including the CG beads whose interactions are to be parameterized, and the results are used to construct an effective potential between the beads. Both approaches have been successfully applied to a number of systems, but potentials derived from all-atom MD simulations carry the added benefit of improved miscibility of all-atom and CG components, which is likely to become increasingly important as mixed all-atom/CG simulations[18–21] become more common.

MD-based parameterization can be carried out in a variety of ways, depending on the scope and intended use of the parameter set in question. Given an all-atom simulation including the components whose interactions are to be parameterized, an effective interaction potential between CG beads can be constructed by attempting to match the forces present between the beads in the all-atom description as a function of distance[22–24,18] or through a process such as Boltzmann inversion[25,26], which is described in more detail in the following sections. Note that although the example given below is for shape-based CG, the same techniques can be applied to determine interactions for residue-based CG models.

Both in the case of MD-based and phenomenological parameterization, the resulting potentials may either be fitted to an existing potential form (for example, the Lennard-Jones potential for nonbonded interactions) or used directly (for example, in the form of an energy/force lookup table). While making use of an existing potential form has long been preferred because it allows the use of existing MD packages without further modification, the use of tabulated potentials allows more control over the exact potential form being used, and is increasingly supported in common MD packages such as DL-POLY and NAMD.

2.2 Reverse coarse graining and resolution switching

Coarse grained MD simulations have proven quite useful for obtaining data on the behavior of systems where the relevant time or length scales (or both) are inaccessible to all-atom MD. However, even heavier use of CG simulations could be made if coarse graining could be used as an accelerator, with atomic detail either maintained in regions of interest or recoverable from snapshots in the CG trajectory. Recent progress has been made along both these fronts recently, in the form of mixed CG-all atom simulations[18] and simulations involving dynamic switching of components between CG and all-atom descriptions[20,21].

The primary new challenges faced in either of these cases lie in deriving accurate potentials for interactions between CG and all-atom components, and in effectively mapping CG conformations to all-atom conformations. The latter challenge is particularly significant both because any given conformation of CG particles can be taken to represent an ensemble of conformations of the corresponding all-atom system (any set of states where the centers of mass of the component atoms for each bead correspond to the CG bead positions), and because switching to the all-atom system will almost certainly cause a change in the energy of the system due to the introduction of new interactions.

Early efforts in switching of scales have focused on building a method allowing true mixed-scale dynamics, either by allowing particles to transition between all-atom and CG representations while passing through a specific region in space[20] or by allowing exchange between low-resolution and high-resolution replicas of a system being simulated in parallel[21]. Outgrowths of these methods will likely be quite useful in the future, although both face the difficulty that deterministically mapping a given CG conformation to an all-atom conformation may be insufficient for more complex beads (such as beads representing an amino acid sidechain or significant fraction thereof) and that the free energy discontinuities experienced during scale-switching may become prohibitively high if a poor initial all-atom conformation is chosen during exchange.

In some cases where a CG model is used to accelerate sampling, there is no need to repeatedly switch between CG and all-atom descriptions; it is sufficient to sample the conformational space of the system using the CG model and then analyze the results in terms of a consistent all-atom model. This is the case, for example, in the studies of nanodiscs presented below, where all-atom conformations had to be extracted from various snapshots of the CG simulation for comparison with experimental data. In this case, it proved sufficient to reverse coarse grain the system by superimposing the all-atom components of the system on the CG structure such that the center of mass of each cluster of atoms is located on the corresponding CG bead, and then minimizing and annealing the resulting all-atom structure with the center of mass of each atom cluster constrained to the bead location (see Fig. 2). This can be conceptually interpreted as sampling the conformational space of the all-atom structure in the region consistent with the CG structure being converted. While this method is far too time-consuming to use when rapid switching of all-atom and CG representations is desired, and does not preserve the dynamic or thermodynamic properties of the CG system, it is sufficient for recovering an all-atom snapshot from a CG simulation, and some conformational sampling scheme similar to that used here is likely to become necessary in resolution exchange for cases where mapping the CG conformation to an all-atom conformation is nontrivial.

2.3 Application to nanodiscs and HDL

High-density lipoproteins (HDL) are lipid-protein particles which function in the body to remove cholesterol from peripheral tissues and return them to the liver for processing. These particles, which occur in a wide variety of shapes and sizes *in vivo*, are known to play an important role in protecting the body from heart disease[27]. HDL particles are known to be composed of a disc-shaped patch of membrane enclosed by two or more copies of apolipoprotein A-I (ApoA-I). In addition to their medical importance, a truncated form of the protein component of HDL particles has recently been used to assemble homogeneous protein-lipid particles known as nanodiscs[28,29], which can incorporate membrane

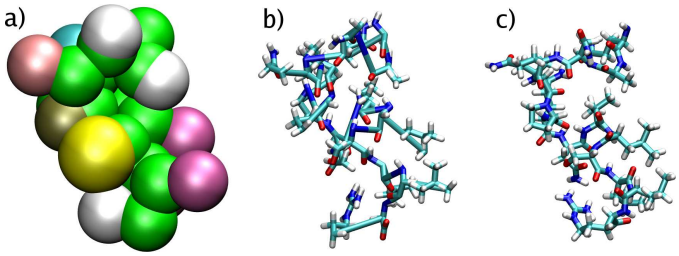


Figure 2. Reverse coarse graining procedure for a short segment of protein. a) Coarse grained representation after simulation. b) Initial all-atom conformation generated through superposition of the all-atom components on the beads. c) Final reverse coarse grained conformation after refinement through constrained dynamics.

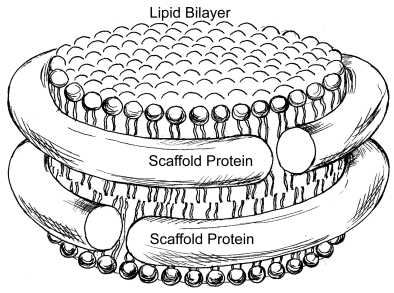


Figure 3. Schematic of the consensus double-belt model for nanodiscs and HDL, showing the position of the ApoA-I proteins wrapped around a lipid disc.

proteins and thus be used to study them in an environment more realistic than micelles or liposomes[30–36].

The conditions needed to cause nanodiscs to assemble around a protein, however, are very dependent on the protein itself, and different conditions are required to efficiently incorporate different proteins[37,38,35]. Obtaining information on the structure and assembly of nanodiscs would thus be useful in the rational design of nanodisc assembly protocols, and would additionally provide data on HDL assembly and characteristics. Unfortunately, no high-resolution structure has been obtained for a complete HDL particle or nanodisc, although a consensus model is emerging for the general layout of the proteins and lipids in the particle, shown in Fig. 3[39–45].

Unfortunately, nanodisc assembly takes place on a timescale of μs to ms, far longer than can be treated using all-atom molecular dynamics simulations. The nature of the type of data sought – relatively coarse data on important stages of nanodisc assembly and factors affecting it – is in principle appropriate for a residue-based CG model. In addition, the fact that hydrophobic interactions and the properties of a lipid patch are the primary features likely to drive the simulation meant that the bulk of the forcefield in this case could be taken from the lipid-water model of Marrink and coworkers[13], a phenomenological model which had shown excellent results in the assembly and physical properties of micelles and bilayers. For the protein component of the system, the bead types of Marrink's forcefield

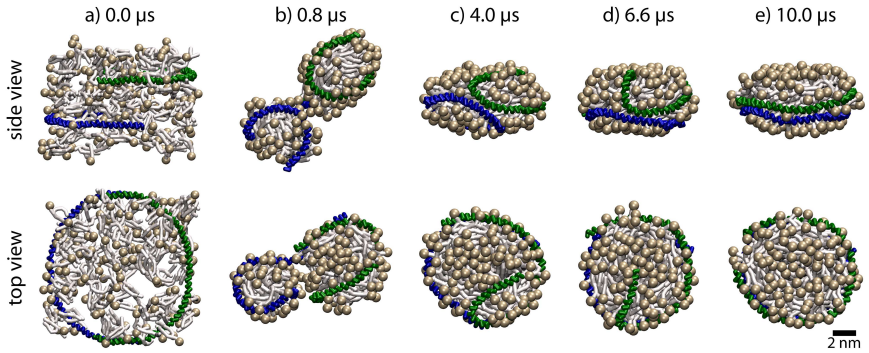


Figure 4. Snapshots from an assembly simulation in which 160 DPPC lipids and two Apo A-1 proteins were assembled from a random mixture over 10 μs. CG water is present in all cases but omitted from images for clarity.

were assigned to protein components according to their properties, with each amino acid residue represented by a backbone bead (the same type for each residue) and a side chain bead[4]. A very similar model was proposed by Bond and coworkers in their simulations of the bacterial membrane protein OmpA[15]. The use of a CG model on the nanodisc provides a factor of 500 speedup compared with all-atom simulations, due to the use of 50 fs timesteps and reduction in number of particles by a factor of 10^4.

Simulation of the components of a single nanodisc beginning from a random mixture with water, over a period of 10 μs, revealed a complete pathway for the assembly of nanodiscs from their components, as shown in Fig. 4. Further simulations from other starting points showed both similar assembly pathways and mechanisms[4,5,46]. Analysis of the energetics of assembly illustrated that it occurs as a three step process. First, nucleation of assembly occurs as the lipids form pseudo-micelles, which are roughly spherical in shape; at this point, the hydrophobic face of the Apo A-I proteins (each of which contains a set of amphipathic α-helices) binds to the pseudo-micelle in a random conformation (Fig. 4b). After this initial aggregation, the proteins reorient along the surface to bring themselves into more favorable contact with each other, eventually forming a series of salt bridges that force the double belt orientation (Fig. 4e) to form.

Although no high-resolution structural data on formed nanodiscs or HDL are available, the assembly mechanism and final structure obtained from CG simulations could still be compared to low resolution information from SAXS studies[47]. Theoretical SAXS curves can be calculated from an all-atom structure using the program CRYSOL[48]; however, obtaining a SAXS curve from CG simulations first requires reverse coarse graining of CG snapshots. Because there was no need to significantly continue the simulations after reverse coarse graining in this case, a fairly simple scheme was used, in which the centers of mass of the all-atom components of each bead were aligned with this bead, and then the system annealed with the center of mass of the components of each bead constrained, allowing the structure to relax while remaining consistent with the CG snapshot. A comparison of the SAXS curve obtained from the assembled CG nanodisc with experimental results is shown in Fig. 5, and a time course of the SAXS curve observed during the CG assembly process in Fig. 6. The excellent agreement between experimental and theoretical

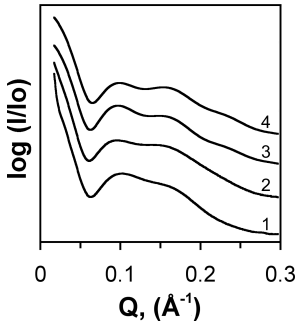


Figure 5. Comparison of SAXS curves between experimental results for DPPC nanodiscs (1), DMPC nanodiscs (2), an ideal all-atom model of a double-belt nanodisc (3), and the final structure from a 10 μs ns CG assembly simulation (4). Note that the curves are separated vertically for clarity.

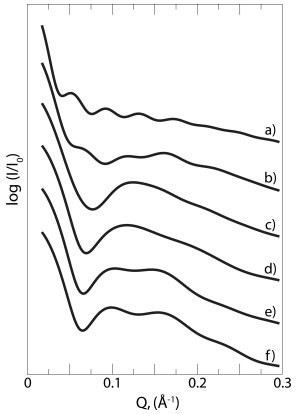


Figure 6. Theoretical SAXS curves obtained through a CG assembly trajectory; timepoints a-f correspond to 0 ns, 150 ns, 850 ns, 1 μs, 4 μs, and 10μs, respectively. Curves are vertically offset from each other for clarity.

results illustrates both the success of the CG model in reproducing the nanodisc assembly process and structure, and the utility of even fairly simple reverse coarse graining methods.

2.4 Application to the BAR domain

BAR domains constitute a ubiquitous type of protein, found in many organisms and performing the function of driving the formation of tubulated and vesiculated membrane structures inside cells[49]. BAR domains involve a conserved protein motif and are involved in a variety of cellular processes including fission of synaptic vesicles, endocytosis, and apoptosis[50]. Structurally, BAR domains form crescent-shaped dimers (see Fig. 7) with a high density of positively charged residues on their concave face. The shape and charge distribution suggest that BAR domains induce membrane curvature by binding to negatively

charged lipids[51,52]. However, the common molecular mechanism underlying membrane sculpting by BAR domains remains largely unknown.

Recently, all-atom simulations[52] have demonstrated that a single BAR domain induces membrane curvature. The all-atom study required a simulation of up to 700,000 atoms on the time scales of $\sim$50 ns. The next demanding question after the discovery of the membrane bending by a single BAR domain is how multiple BAR domains work together to bend membranes. All-atom simulations of this process are too challenging at present, since one would have to consider millions of atoms in each simulation. However, the residue-based CG method appears to be a good option for this application, and, thus, we have performed CG simulations of systems with multiple BAR domains, in order to determine how the cooperative interaction of the latter with the membrane induces global membrane curvature.

The residue-based CG model[5,4] described above is ideally suited to describe the membrane remodeling by BAR domains since it has demonstrated its power before on the tasks where lipids assemble, disassemble, and re-shape membranes[5,4,13]. The only difficulty is that the residue-based protein CG model has not been developed to work for proteins of arbitrary shapes. In particular, the model has not been designed to maintain tertiary structure of proteins, which is determined by the protection of hydrophobic side groups in the protein amino acid sequence from solvent (well described by the residue-based CG mode), but also, to a large extent, by atomic level interactions that the residue-based CG model does not capture. Indeed, when the model was applied to the BAR domain, the tertiary structure was not preserved. Accordingly, we added harmonic bonds and angles connecting protein beads that conserve protein shape and flexibility. A minimal set of bonds and angles was selected for this purpose. The strength of these bonds and angles was chosen to reproduce the tertiary structure flexibility as observed in the all-atom simulations. As a result, the protein was not heavily constrained, but the tertiary structure (the BAR domain's crescent shape) was maintained well. This feature has been implemented through a NAMD[53] functionality that allows one to add extra bonded interactions to simulations.

In our previous residue-based CG simulations[5,4,13], a relative dielectric constant ε of 20 was employed. In the case of the BAR domain simulations we chose $\varepsilon = 1$. Such low ε-value is necessary for membrane curvature to be induced by BAR domains, which is driven by short-range electrostatics, when charged groups from the protein's concave surface interact at close range with charged lipid heads. Interactions at larger distances should be screened by water requiring in principle higher values of ε. However, the electrostatic interactions at large distances appear to be relatively weak in the present case such that $\varepsilon = 1$ has no adverse effect on long-range electrostatics in case of the BAR domain simulations.

The rather rough CG model of the BAR domain and lipid membrane, described above, has been applied to study the behavior of multiple BAR domains[54], as shown in Fig. 7. The all-atom simulations with a single BAR domain[52], from other groups as well as our own, have been reproduced well by the residue-based CG simulations (not shown), in terms of both membrane curvature and protein structure. Six BAR domains interacting with a patch of membrane were then simulated. Two rows of three BAR domains each were placed in parallel (shifted with respect to each other) on top of a planar membrane, composed of electrostatically neutral DOPC lipids mixed with negatively charged DOPS lipids (30 % DOPS). BAR domains produced a global bending mode[54], exhibiting a radius of curvature

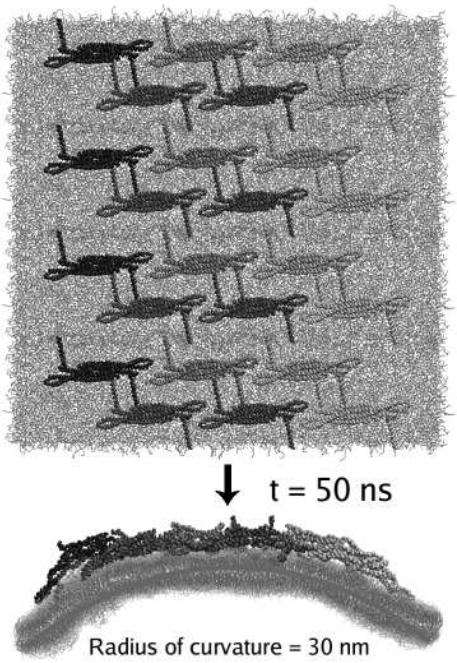


Figure 7. Membrane curvature induced by BAR domains. Upper panel: top view of the initial arrangement (four periodic cells along the vertical axis); lower panel: side view after 50 ns.

of 30 nm within 50 ns (comparable to experimental values for the curvature[51]). This result suggests how BAR domains generate quickly membrane curvature, as possibly occurs in cells during the formation of sub-cellular membrane structures[50].

3 Shape-Based Coarse Graining

The shape-based CG[55,56] method offers a higher degree of coarse graining than the residue-based method, but a the price that the biopolymers described are restricted in their motion to elastic vibration around a morphology. The method is available through the molecular visualization software VMD[a][57].

3.1 Selection of bead arrangement and potentials

Biomolecules, and proteins in particular, assume a variety of shapes, often featuring both compact domains and elongated tails, the compact regions and tails often being equally important. To our knowledge, all existing CG methods assign CG beads to represent a fixed group of atoms, but this is not efficient for the coarse graining of molecules with complex shapes, because with such an approach either the tails are misrepresented, or

[a] http://www.ks.uiuc.edu/Research/vmd/

too many CG beads are used for the compact domains. With the shape-based CG, one addresses the task of representing shapes with as few CG beads as possible by so-called topology conserving maps[58].

Consider a molecule consisting of N_a atoms with coordinates $\mathbf{r}_n$ and masses m_n, $n = 1, 2, ..., N_a$. One seeks to reproduce the shape of the molecule with N CG beads. The mass distribution $p_n = m_n/M$ ($M = \sum_n m_n$) is used as a target probability distribution for the evolving map. CG beads are assigned their initial positions randomly; then, the beads are considered as nodes of a network[58], on which S adaptation steps are performed. At each step the following procedures are carried out. First, the n-th atom is chosen randomly, according to the probability distribution p_n; its coordinates $\mathbf{r}_n = \mathbf{v}$ are used to adapt the neural network (see Eq. 1). Second, for each CG bead i ($i = 1, 2, ..., N$), one determines the number k_i of CG beads j, obeying the condition $|\mathbf{v} - \mathbf{R}_j| < |\mathbf{v} - \mathbf{R}_i|$, where $\mathbf{R}_j$ is the position of the j-th bead. Third, positions of the beads are updated ($i = 1, 2, ..., N$), according to the rule

$$\mathbf{R}_i^{new} = \mathbf{R}_i^{old} + \epsilon e^{-k_i/\lambda}(\mathbf{v} - \mathbf{R}_i^{old}). \tag{1}$$

Parameters ϵ and λ are adapted at each step according to the functional form $f_s = f_0(f_S/f_0)^{s/S}$, where s is the current step, $\lambda_0 = 0.2N$, $\lambda_S = 0.01$, $\epsilon_0 = 0.3$, and $\epsilon_S = 0.05$. We use $S = 200N$; typical adaptation steps are shown in Fig. 8. Once beads are placed, an all-atom "domain" is found for each bead (the domain includes all atoms closer to this bead than to any other bead). The total mass and charge of a domain is assigned to the respective bead. Since the shape of a molecule is reproduced by this CG model, the method is termed shape-based CG. The molecular graphics program VMD[57], through its shape-based CG plugin[b], can also build CG models from volumetric data, such as density maps obtained from cryo-electron microscopy.

Currently, two ways of establishing bonds between CG beads are implemented. In one case, a bond is established if the distance between two beads is below a cutoff distance (chosen by the researcher). Another possibility is to establish a bond between two CG beads if their respective all-atom domains are connected by protein or nucleic backbone trace; in the latter case, the topology of the molecular polymeric chain is reproduced better.

Interactions between beads are described by a CHARMM-like force-field[59], i.e., bonded interactions are represented by harmonic bond and angle potentials (no dihedral potentials). The non-bonded potentials include 6-12 Lennard-Jones (LJ) and Coulomb terms

$$V = \sum_{bonds\, i} \frac{K_i}{2}(R_i - L_i)^2 + \sum_{angles\, k} \frac{M_k}{2}(\theta_k - \Theta_k)^2 + \sum_{m,n} 4\epsilon_{mn}\left[\left(\frac{\sigma_{mn}}{r_{mn}}\right)^{12} - \left(\frac{\sigma_{mn}}{r_{mn}}\right)^{6}\right] + \sum_{m,n} \frac{q_m q_n}{4\pi\varepsilon\varepsilon_0 r_{mn}}, \tag{2}$$

where R_i and θ_k are the distance and angle for bond i and angle k, K_i and M_k are the force constants, L_i and Θ_k are the equilibrium bond length and angle; r_{mn} is the distance between beads m and n, ϵ_{mn} and σ_{mn} are the LJ parameters, q_m is the charge of the mth bead, and the sum over m and n runs over all pairs of CG beads. The constant ε_0 is the vacuum dielectric permittivity; ε is a relative dielectric constant.

[b] http://www.ks.uiuc.edu/Research/vmd/plugins/cgtools/

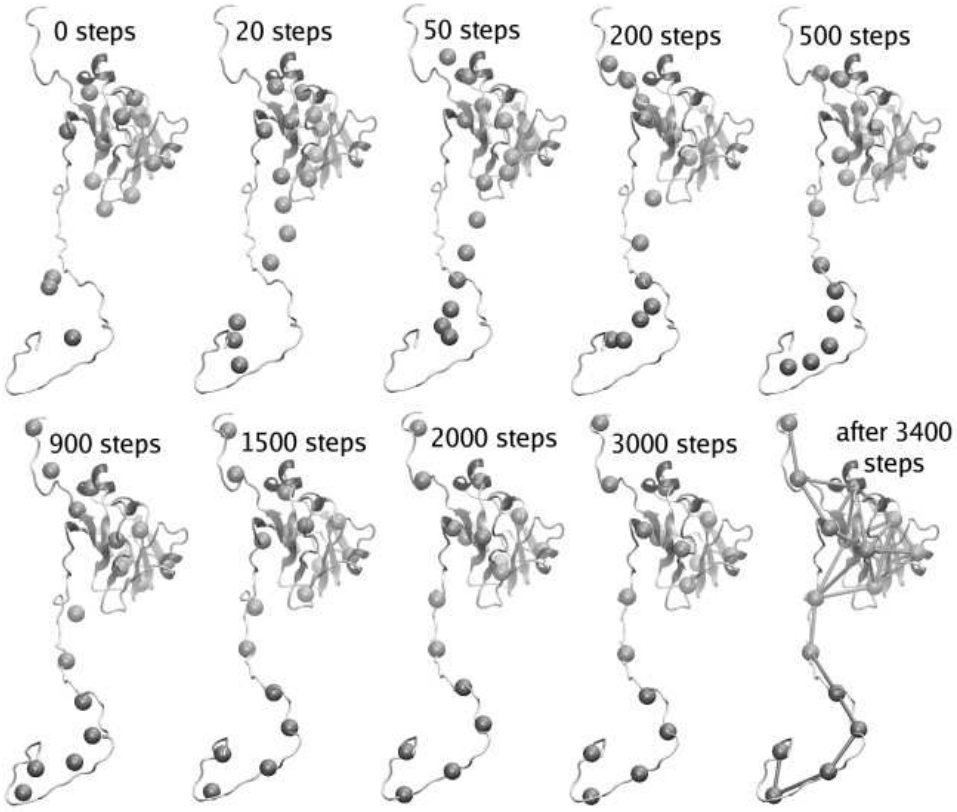


Figure 8. Shape-based coarse graining algorithm assigning CG beads. The CG beads (spheres) are the nodes of the network; their positions are updated throughout the learning steps (3400 steps for 17 beads in this example). As a result, the shape of a protein (here, the capsid unit protein of brome mosaic virus) is reproduced with a small number of beads (chosen prior to starting the algorithm). After the assignment converged, the beads are connected by bonds. The algorithm is of a neural network type described in[58].

Bonded parameters K_i, L_i, ϵ_{mn}, etc., can be extracted from all-atom MD simulations of the considered system. For each CG bond and angle, one follows the distances between the centers of mass of corresponding atomic domains; CG force-field parameters are chosen so that in the CG simulation of a protein unit, the mean distances (angles) and respective root mean square deviations (rmsd) reproduce those found in an all-atom simulation. This procedure can be illustrated by the simple example of a one-dimensional harmonic oscillator, with a particle moving along the x coordinate in the potential $V(x) = f(x - x_0)^2/2$. With the system in equilibrium at temperature T, the average position $\langle x \rangle$ is equal to x_0, and the rmsd is given by $k_B T/f$ (k_B is the Boltzmann constant). Using an MD simulation, one can compute $\langle x \rangle$ and the rmsd, thus obtaining x_0 and f.

In all-atom simulations, LJ radius σ_{mn} for a pair m, n is usually approximated by $\sigma_{mn} = (\sigma_m + \sigma_n)/2$, where σ_m is the LJ radius of the m-th atom. We use the same

approach for CG beads; σ_m for the m-th bead is calculated as the radius of gyration of its all-atom domain, increased by 2 Å (an average LJ radius of an atom in the CHARMM force-field). The LJ well depth ϵ_{mn} is set to a uniform value for all pairs $m-n$; usually, we used $\epsilon_{mn} = 4$ kcal/mol. This choice for σ_{mn} and ϵ_{mn} was supported by all-atom simulations of pairs of protein segments about 500 atoms each (roughly representing a single CG bead in one of our applications). Several such simulations were performed, for about 10 ns each. The effective potential of interaction between two segments was obtained for every pair using the Boltzmann inversion method[25,26]: assuming that the distribution of the distance between the segments x is given by $\rho(x) = e^{-V(x)/(k_BT)}$, where $V(x)$ is the potential, one computes $\rho(x)$ from the simulation and finds the potential as $V(x) = -k_BT\ln[\rho(x)] + const$. The potentials computed from all-atom simulations were similar to a LJ potential in shape, and for each pair the well depth was about 4 kcal/mol; the LJ radius was well represented using the procedure (radius of gyration + 2 Å) described above[56].

An effect of the solvent model is modeled implicitly, by reproducing three basic features of water, namely, viscosity, fluctuations due to Brownian motion, and dielectric permittivity. The relative dielectric constant ε is set to 80 everywhere (the experimental value for liquid water). Frictional and fluctuating forces are introduced through the Langevin equation that describes the time evolution of the CG system for each bead

$$m\ddot{\mathbf{r}} = \mathbf{F} - m\gamma\dot{\mathbf{r}} + \chi\psi(t). \tag{3}$$

Here, $\mathbf{r}$ is the position of the bead, $\mathbf{F}$ is the force acting on the bead from other beads in the system, γ is a damping coefficient, $\psi(t)$ is a univariate Gaussian random process, and χ is related to the frictional forces through the fluctuation-dissipation theorem, $\chi = \sqrt{2\gamma k_BT/m}$, with m being the bead's mass. With $\mathbf{F} = 0$, Eq. 3 describes free diffusion, where γ is related to the diffusion constant D, $D = k_BT/(m\gamma)$. In principle, γ can be computed from all-atom simulations by calculating D for the molecule under study (although the force fields used in such simulations might be not good enough to reproduce the water viscosity), but a much better approach is to use an experimental value of D if available, e.g., D for a molecule of similar size. Contrary to the extraction of D from all-atom simulation, which is often difficult due to insufficient sampling, γ can be easily tuned in CG simulations to give the appropriate value of D for a given molecule, since one achieves sampling for the center of mass displacements much faster in CG simulations than in all-atom simulations. Based on estimates from the all-atom simulations and experimental data for various proteins, the appropriate values of γ for 500 atoms per CG bead should be in the range 3-15 ps^{-1}.

The dynamics of the CG system is realized through MD simulations using NAMD[53]. For the case of 500 atoms per CG bead the coarse graining allows one to simulate systems 500 times larger than possible in all-atom representation. As water often accounts for $\sim 80\%$ of atoms in biomolecular simulations, and since the solvent is treated implicitely, the real gain is even higher, typically 2,000-3,000 times. Due to slower motions of CG beads in comparison with atoms, one can use a time step of $\sim$500 fs to integrate the equations of motion, instead of 1 fs common for all-atom simulations. As a result, the shape-based CG with a typical ratio of 500 atoms per bead allows one to simulate dynamics of micrometer-sized objects on time scales of 100 μs using just 1-3 processors, while all-atom simulations even with 1,000 processors are limited now to $\sim$20 nm in size and 100 ns in time. Of

course, this gain comes at the price of limited resolution.

3.2 Application to Structural Dynamics of Viruses

Shape-based CG was successfully applied to study the structural dynamics of viruses. A virus[60,61] is a macromolecular complex, normally 10-100 nm across, consisting of a genome enclosed in a protein coat (capsid); usually, the capsid is a symmetric assembly, often an icosahedron, formed by multiple copies of a few proteins. Other accessory molecules can be contained inside the capsid; additional proteins and a lipid bilayer envelope are also found on the surface of some viruses. The viral replication cycle starts with the delivery of the viral genome into a host cell, a step usually involving capsid disintegration. Then, the host cell replicates the viral genome and produces viral proteins, often at the cost of reducing the cell's normal functionality. Finally, the newly produced parts of the virus assemble into viral particles and leave the host cell, which is usually destroyed as a result. Outside of the host cell a viral particle has to be stable and relatively rigid to protect the genome, but it also has to become unstable when virulence factors need to be released into the host cell. In order to determine the stability of viral capsids and transitions between stable and unstable structures, we performed MD simulations of several viruses, both in all-atom[62] and CG representations[55].

Employing the shape-based CG method[55], we were able to study large viral capsids (up to 75 nm in diameter, see Fig. 9) on 1.5-25 μs time scales. Most of the simulations were performed on a single processor, but parallel simulations on up to 48 processors were also carried out; the latter exhibited good parallel scaling similar to that of all-atom simulations with NAMD[53].

First[55], we performed CG simulations of satellite tobacco mosaic virus (STMV), found in good agreement with previous all-atom simulations[62]. STMV is one of the smallest and simplest viruses, only 17 nm in diameter (Fig. 9), yet, to describe it using all-atom simulations required dealing with a one-million-atom system. MD simulations on the complete STMV showed that it is perfectly stable on a time scale of 10 ns. The STMV capsid without genome, in contrast, was unstable, showing a remarkable collapse over the first 5-10 ns of simulation. The CG simulation of STMV reproduced the patterns and timescales of the collapse observed for the STMV capsid in all-atom simulations. For both complete STMV and the capsid alone, several other quantities computed in CG simulations, such as the average capsid radius, were within a few Å from those in the all-atom study.

CG simulations of capsids of several more viruses were then carried out (Fig. 9), of the satellite panicum mosaic virus (SPMV), the satellite tobacco necrosis virus (STNV), the brome mosaic virus (BMV), the poliovirus, the bacteriophage ϕX174, and reovirus. In CG simulations, the empty capsids of STMV, SPMV, and STNV collapsed. The reovirus core, the bacteriophage ϕX174 procapsid, and the poliovirus capsid were stable, and indeed, it is known experimentally that these are stable even without their respective genetic material. For BMV, empty capsids have been observed experimentally, while a cleavage of the N-terminal tails of the unit proteins makes the capsid unstable[63]. In agreement with that, the BMV capsid was stable in our simulations, although very flexible, but when the N-terminal tails were removed, the capsid collapsed.

Thus, results of CG simulations agree with all-atom studies and experimental data, where available. The simulations provide also new quantitative information about viral

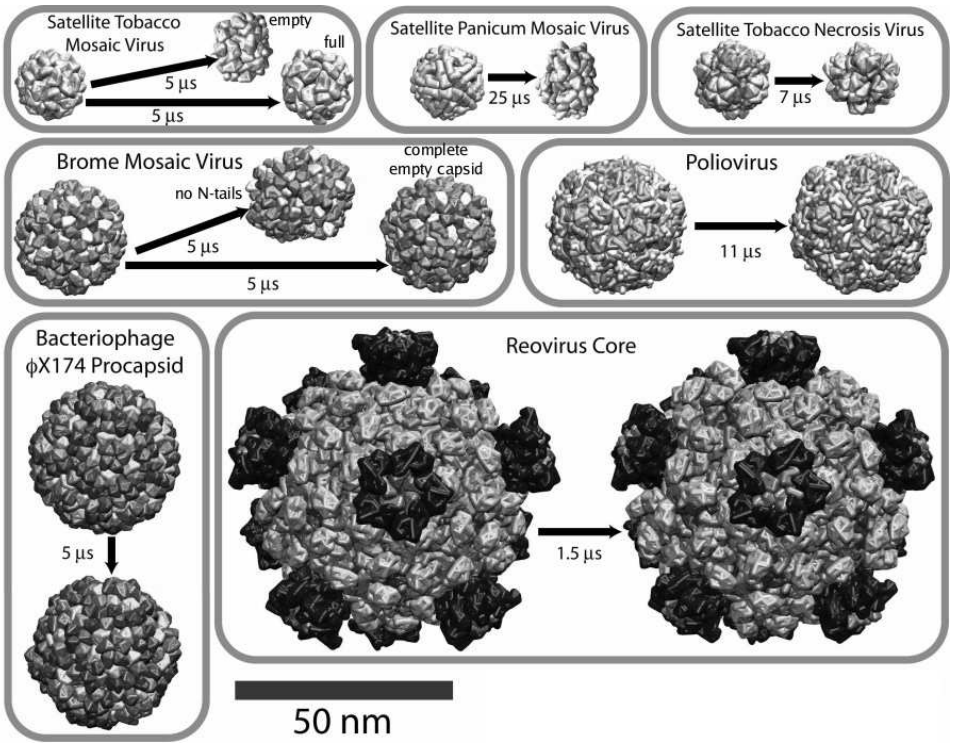


Figure 9. CG simulations of viral capsids. The initial and final structures for each simulation are shown (all particles are drawn to scale). The ratio of ∼200 atoms per CG particle is used. All capsids are simulated without gene content, i.e., empty, except in case of satellite tobacco mosaic virus, in which case both empty and full capsids were simulated. From[55].

dynamics. Perhaps the main finding in this regard is that some of the capsids (STMV, SPMV, and STNV) cannot maintain their structural integrity in the absence of the genome. This suggests a specific self-assembly pathway for these viruses: it must be the RNA, and not the protein, which nucleates assembly of the complete virus. Probably, the RNA forms a spherical particle, and then capsid proteins attach to its surface. It is known for some viruses that they assemble "capsid first"[61], the genome being pulled into the pre-formed capsid. Our simulations and emerging experimental evidence[63,64] suggest that this might be different for some viruses. Related to what determines the stability, we found that the stability and flexibility of viral capsids is closely correlated with the strength of interactions between capsid subunits. Larger capsids, such as the reovirus core, have proteins that intricately intertwine with each other, featuring even a "thread and needle" arrangement. For STMV, SPMV, and STNV, unit proteins only touch each other by the edges. With more contacts between the protein units, a capsid has more hydrogen bonds and salt bridges per area unit (reflected in the CG model by generalized non-bonded LJ and Coulomb forces), and the frictional force between the capsid faces rises. These factors enhance the stability. Our simulations suggest that viruses like STMV, SPMV, and STNV have relatively few contacts between the capsid subunits and only their genomes render the capsids stable.

3.3 Application to the bacterial flagellum

The shape-based CG method has recently been applied also to study the molecular basis of bacterial swimming. Many types of bacteria propel themselves through liquid media using whip-like structures known as flagella. The bacterial flagellum is a huge (several μm long, $\sim$20 nm wide), multiprotein assembly built of three domains: a basal body, fixed in the cell body below the outer membrane and acting as a motor; a filament, which grows out of the cell, making up the bulk of the length of the flagellum and interacting with solvent to propel the bacterium; and a hook, connecting basal body and filament and acting as a joint transmitting the torque from the former to the latter. Depending on the direction of the torque applied by the basal body, the filament assumes different helical shapes. Under counter-clockwise rotation (as viewed from the exterior of the cell), several flagella form a single helical bundle which propels the cell along a straight line (running mode)[65]. Under clockwise rotation, the individual flagella dissociate from the bundle and form separate right-handed helices, causing the cell to tumble. Varying the duration of running and tumbling, bacteria can move up or down a gradient of an attractant or repellent by a biased random walk.

One of the unresolved questions about the flagellum is how the reversal of torque applied by the motor results in a switching between the helical shapes of the flagellar filament. This switching is a result of polymorphic transitions in the filament, when individual protein units slide against each other[66], but its molecular mechanism remains poorly understood. Trying to answer this question, we performed CG MD studies of the flagellar filament[56], which is formed by thousands of copies of a single protein, flagellin. Flagellin was coarse grained with $\sim$500 atoms per CG bead (vs. $\sim$200 for viruses), as shown in Fig. 10. Segments of the filament (1,100 flagellin unit, or $\sim$0.5 μm long) were rotated clockwise and counter-clockwise, with a constant rotation speed one turn in 10 μs applied to 33 protein units at the bottom of the segment. The simulations covered 30 μs each.

The filament is built by the helical arrangement of flagellin units, eleven per turn. A thread of units each separated by one turn is called "protofilament" (see Fig. 10); eleven protofilaments comprise the filament. In the CG simulations, the filament segments remained stable when rotated, but protofilaments rearranged dramatically (though it must be noticed that the toque applied to the model flagellum exceeded by far the one arising under native conditions). In the straight filament, which was the starting structure, the protofilaments form a right-handed helix with large helical period. When the torque is applied counterclockwise (as viewed from the base to the tip), the protofilaments remain arranged in right-handed helices, but the pitch of the helices rises; when the torque is opposite, the helices become left-handed. The filament also forms a helix as a whole. For the rotation corresponding to the running mode, the filament forms a left-handed helix, whereas for the tumbling mode it becomes a right-handed helix. The same difference in handedness between these helices is found in living bacteria[67].

Running and tumbling modes of bacterial swimming are determined by structural transitions in the flagellar filament, depending on the direction of the applied torque. Clearly, interactions between protein units play an important role in enabling this transition. However, flagella act in solvent (water), and, curiously, the role of the solvent has not been analyzed much before. Using the simple description by means of Eq. 3, where viscosity is governed by a single parameter γ, one can investigate the effect of solvent[56]. It was found that without friction due to solvent, flagella rotate as a rigid body, i.e., the mutual positions

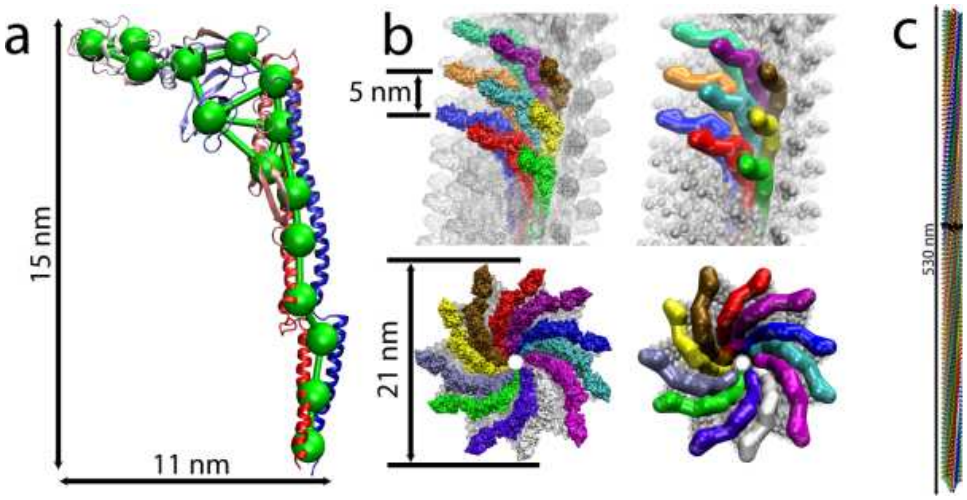


Figure 10. Coarse graining of the flagellar filament. Unit proteins are represented by 15 CG beads (a). In (b), the flagellar filament viewed from the side and from the top is shown in all-atom (left) and CG (right) representations. A filament segment (1,100 monomers) is shown in CG representation in (c). A single helix turn of eleven unit proteins is highlighted in black.

of monomers are frozen, both for running and tumbling mode. With the solvent's friction present, the protofilaments rearrange as explained above, in agreement with structural changes in the flagellum suggested by experimental studies. Thus, the solvent (friction) plays a crucial role in the switching between the arrangements of protofilaments and, consequently, in producing supercoiling along the entire filament, or running and tumbling modes of motion.

4 Future Applications of Coarse Graining

Due to growing interest in large biomolecules and systems biology, coarse grained simulations have grown increasingly common over the past few years as a means of accessing time and size scales that cannot be reached with all-atom molecular dynamics. Recent advances such as more reliable forcefields for residue-based coarse graining[17,68], mixed CG and all-atom simulations[18,20], and low resolution shape-based CG models[55,56] have improved the accuracy, flexibility, and potential scope of CG simulations. Since, however, coarse grained simulations will never offer the same level of accuracy as all-atom simulations, it seems likely that CG simulations will naturally evolve in directions allowing closer links to atomistic descriptions. Both the aforementioned techniques of dynamic changes of scale and mixing CG and all-atom descriptions serve as useful and distinct models for how this can be accomplished, with the former using coarse graining as an accelerant to improve sampling and then using all-atom simulations to flesh out the details of the sampled states, and the latter allowing less important parts of a system (such as bulk solvent) to be treated with a lower resolution than the regions of interest.

The utility of further development and application of these techniques can be seen,

for example, through consideration of the bacterial flagellum. Coarse grained simulations have been used to investigate both the large-scale behavior of the flagellar filament during supercoiling[56] and solvent dynamics around the supercoiled flagellum[69]; at the same time, large-scale all-atom simulations have offered a potential atomic-scale mechanism for differential supercoiling[70]. The remaining challenge for theory is to fully link the CG and atomistic descriptions to provide a coherent and fully testable model for filament supercoiling; the most likely path for developing such a model is to use rotation of a shape-based CG filament to develop an ensemble of conformations at different points along the flagellum, which can then be simulated and perturbed in an all-atom representation to understand what interactions and structural transitions are important for the supercoiling process. A similar scale-switching approach could be applied to other systems, including viral capsids (allowing the study of assembly intermediates obtained from shape-based coarse graining).

The shape-based CG methods should be further developed in a few important directions. Our present shape-based CG methodology[55,56] allows one to simulate proteins. Despite initial successes, the protein model remains relatively rough and needs to be further refined, in particular with respect to the interaction potentials employed. These potentials can be improved using systematic all-atom parameterizing simulations for target systems. The same is true for the solvent model, which should be further developed along the lines of a true implicit solvent model, such as the generalized Born approach[71–73]. The CG method should also be extended to biomolecules other than proteins; to that end, we have recently started the development of a shape-based CG membrane model[54]. In this model, each leaflet of a lipid bilayer is represented by a collection of two-bead "molecules" (two beads connected by a spring), held together by non-bonded interactions tuned to mimic the bilayer stability, thickness, and area per lipid. This approach is somewhat similar to previous attempts of CG membrane simulations, such as in[74]. However, in our model each two-bead "molecule" represents a patch of a leaflet (not necessarily an integer number of lipid molecules), rather than a single lipid. Using the model, we have been able to simulate bilayer self-assembly and reproduce the results of all-atom and residue-based CG simulations of BAR domains (see above); much larger BAR domain simulations using the new model are under way. The shape-based CG model describing proteins and lipids will be very useful for simulations of sub-cellular processes, where multiple proteins interact with each other and with cellular membranes on long timescales.

Future residue-based CG simulations of nanodiscs will continue to further our understanding of HDL assembly and maturation, as well as aiding in the use of synthetic nanodiscs as protein scaffolds. HDL particles acting *in vivo* absorb esterified cholesterol for transport[27]; understanding the structural transitions involved in this process will be a key step in the overall goal of characterizing HDL function. This absorption process can be studied through residue-based CG simulations designed to observe how the structure of a nanodisc adjusts to the presence of esterified cholesterol. Ongoing simulations of nanodiscs will also be used to refine reverse coarse graining methods for residue based CG models to move from the snapshot-only reversal described above to a thermodynamically correct method for changing from all-atom to residue-based CG models.

The continued development and application of coarse graining, along with ongoing improvements in generally available computational resources, promises to enable biomolecular simulations to treat many systems which were previously inaccessible. The increasing application of all-atom and CG simulations to the same system should greatly increase

the impact of coarse graining by allowing the acceleration of coarse grained simulation to be obtained without sacrificing atomic detail. At the same time, pure CG simulations also continue to be useful for understanding the behavior of large systems over very long timescales, and this utility will only increase with continuing improvements to CG potentials.

References

1. Markos A Katsoulakis, Andrew J Majda, and Dionisios G Vlachos, *Coarse-grained stochastic processes for microscopic lattice systems.*, PNAS, **100**, no. 3, 782–787, Feb 2003.
2. Kumara Sastry, D. D. Johnson, David E. Goldberg, and Pascal Bellon, *Genetic programming for multitimescale modeling*, Phys. Rev. B, **72**, no. 8, 085438–9, Aug. 2005.
3. Ch. Schütte, A. Fischer, W. Hiosinga, and P. Deuflhard, *A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo*, Journal of Computational Physics, **151**, 146–168, 1999.
4. Amy Y. Shih, Anton Arkhipov, Peter L. Freddolino, and Klaus Schulten, *Coarse grained protein-lipid model with application to lipoprotein particles*, J. Phys. Chem. B, **110**, 3674–3684, 2006.
5. Amy Y. Shih, Peter L. Freddolino, Anton Arkhipov, and Klaus Schulten, *Assembly of Lipoprotein Particles Revealed by Coarse-Grained Molecular Dynamics Simulations*, J. Struct. Biol., **157**, 579–592, 2007.
6. Andrew R. Leach, *Molecular Modelling, Principles and Applications*, Addison Wesley Longman Limited, Essex, 1996.
7. J C Shelley, M Y Shelley, R C Reeder, S Bandyopadhyay, P B Moore, and M L Klein, *Simulations of phospholipids using a coarse grain model*, J. Phys. Chem. B, **105**, 9785–9792, 2001.
8. M. J. Stevens, J H Hoh, and T B Woolf, *Insights into the Molecular Mechanism of Membrane Fusion from Simulations: Evidence for the Association of Splayer Tails*, Phys. Rev. Lett., **91**, 188102, 2003.
9. Mark J. Stevens, *Coarse-grained simulations of lipid bilayers*, J. Chem. Phys., **121**, 11942–11948, 2004.
10. Steve O. Nielsen, Carlos F. Lopez, Goundla Srinivas, and Michael L. Klein, *Coarse grain models and the computer simulation of soft materials*, J. Phys.: Condens. Matter, **16**, no. R481–R512, 2004.
11. Jens Erik Nielsen and J. Andrew McCammon, *On the evaluation and optimization of protein X-ray structures for pK_a calculations*, Prot. Sci., **12**, 313–326, 2003.
12. S. J. Marrink and A. E. Mark, *Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles*, J. Am. Chem. Soc., **125**, no. 49, 15233–42, 2003.
13. Siewert J. Marrink, Alex H. de Vries, and Alan E. Mark, *Coarse Grained Model for Semiquantitative Lipid Simulations*, J. Phys. Chem. B, **108**, 750–760, 2004.
14. Siewert J. Marrink, Jelger Risselada, and Alan E. Mark, *Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model*, Chem. Phys. of Lipids, **135**, no. 2, 223–244, 2005.

15. Peter J. Bond and Mark S. P. Sansom, *Insertion and Assembly of Membrane Proteins via Simulation*, J. Am. Chem. Soc., **128**, 2697–2704, 2006.
16. Payel Das, Silvina Matysiak, and Cecilia Clementi, *Balancing energy and entropy: A minimalist model for the characterization of protein folding landscapes*, Proc. Natl. Acad. Sci. USA, **102**, no. 29, 10141–10146, 2005.
17. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, *The MARTINI forcefield: coarse grained model for biomolecular simulations*, J. Phys. Chem. B, **111**, 7812–7824, 2007.
18. Qiang Shi, Sergei Izvekov, and Gregory A. Voth, *Mixed Atomistic and Coarse-Grained Molecular Dynamics: Simulation of a Membrane-Bound Ion Channel*, J. Phys. Chem. B, **110**, no. 31, 15045–15048, 2006.
19. Matej Praprotnik, Luigi Delle Site, and Kurt Kremer, *Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly*, J. Chem. Phys., **123**, 224106–1–224106–14, 2005.
20. Matej Praprotnik, Luigi Delle Site, and Kurt Kremer, *Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids*, Phys. Rev. E, **73**, 066701, 2006.
21. Edward Lyman, F. Marty Ytreberg, and Daniel M Zuckerman, *Resolution exchange simulation.*, Phys Rev Lett, **96**, no. 2, 028105, Jan 2006.
22. Sergei Izvekov and Gregory A Voth, *Multiscale coarse graining of liquid-state systems.*, J. Chem. Phys., **123**, 134105, 2005.
23. Sergei Izvekov and Gregory A Voth, *A multiscale coarse-graining method for biomolecular systems.*, J Phys Chem B, **109**, no. 7, 2469–2473, Feb 2005.
24. Sergei Izvekov and Gregory A. Voth, *Multiscale Coarse-Graining of Mixed Phospholipid/Cholesterol Bilayers*, J. Chem. Theory Comput., **2**, 637–648, 2006.
25. Dirk Reith, Mathias Pütz, and Florian Müller-Plathe, *Deriving effective mesoscale potentials from atomistic simulations*, J. Comp. Chem., **24**, 1624–1636, 2003.
26. Valentina Tozzini and Andrew McCammon, *A coarse grained model for the dynamics of flap opening in HIV-1 protease*, Chem. Phys. Lett., **413**, 123–128, 2005.
27. Minghan Wang and Michael R Briggs, *HDL: The metabolism, function, and therapeutic importance.*, Chem. Rev., **104**, 119–137, 2004.
28. Timothy H. Bayburt, Yelena V. Grinkova, and Stephen G. Sligar, *Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins*, Nano Letters, **2**, 853–856, 2002.
29. S G Sligar, *Finding a single-molecule solution for membrane proteins*, Biochem. Biophys. Res. Commun., **312**, 115–119, 2003.
30. Annela M Seddon, Paul Curnow, and Paula J Booth, *Membrane proteins, lipids and detergents: not just a soap opera*, Biochim. Biophys. Acta, **1666**, 105–117, 2004.
31. Dmitri R Davydov, Harshica Fernando, Bradley J Baas, Stephen G Sligar, and James R Halpert, *Kinetics of dithionite-dependent reduction of cytochrome P450 3A4: heterogeneity of the enzyme caused by its oligomerization.*, Biochemistry, **44**, 13902–13913, 2005.
32. Bradley J Baas, Ilia G Denisov, and Stephen G Sligar, *Homotropic cooperativity of monomeric cytochrome P450 3A4 in a nanoscale native bilayer environment.*, Arch. Biochem. Biophys., **430**, 218–228, 2004.
33. Hui Duan, Natanya R Civjan, Stephen G Sligar, and Mary A Schuler, *Co-*

incorporation of heterologously expressed Arabidopsis cytochrome P450 and P450 reductase into soluble nanoscale lipid bilayers., Arch. Biochem. Biophys., **424**, 141–153, 2004.
34. Natanya R Civjan, Timothy H Bayburt, Mary A Schuler, and Stephen G Sligar, *Direct solubilization of heterologously expressed membrane proteins by incorporation into nanoscale lipid bilayers.*, Biotechniques, **35**, 556–560, 562–563, 2003.
35. T Boldog, S Grimme, M Li, S G Sligar, and G L Hazelbauer, *Nanodiscs separate chemoreceptor oligomeric states and reveal their signaling properties*, Proc. Natl. Acad. Sci. USA, **103**, 11509–11514, 2006.
36. Amy Y. Shih, Ilia G. Denisov, James C. Phillips, Stephen G. Sligar, and Klaus Schulten, *Molecular Dynamics Simulations of Discoidal Bilayers Assembled from Truncated Human Lipoproteins*, Biophys. J., **88**, 548–556, 2005.
37. I G Denisov, Y V Grinkova, A A Lazarides, and S G Sligar, *Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size.*, J. Am. Chem. Soc., **126**, 3477–3487, 2004.
38. Timothy H Bayburt, Yelena V Grinkova, and Stephen G Sligar, *Assembly of single bacteriorhodopsin trimers in bilayer nanodiscs.*, Arch. Biochem. Biophys., **450**, 215–222, 2006.
39. V Koppaka, L Silvestro, J A Engler, C G Brouillette, and P H Axelsen, *The structure of human lipoprotein A-I. Evidence for the "belt" model.*, J. Biol. Chem., **274**, 14541–14544, 1999.
40. S E Panagotopulos, E M Horace, J N Maiorano, and W S Davidson, *Apolipoprotein A-I adopts a belt-like orientation in reconstituted high density lipoproteins.*, J. Biol. Chem., **276**, 42965–42970, 2001.
41. H Li, D S Lyles, M J Thomas, W Pan, and M G Sorci-Thomas, *Structural determination of lipid-bound ApoA-I using fluorescence resonance energy transfer.*, J. Biol. Chem., **275**, 37048–37054, 2000.
42. M A Tricerri, A K Behling Agree, S A Sanchez, J Bronski, and A Jonas, *Arrangement of apolipoprotein A-I in reconstituted high-density lipoprotein disks: an alternative model based on fluorescence resonance energy transfer experiments.*, Biochemistry, **40**, 5065–5074, 2001.
43. R A G D Silva, George M Hilliard, Ling Li, Jere P Segrest, and W Sean Davidson, *A mass spectrometric determination of the conformation of dimeric apolipoprotein A-I in discoidal high density lipoproteins.*, Biochemistry, **44**, 8600–8607, 2005.
44. Y Li, A Z Kijac, S G Sligar, and C M Rienstra, *Structural Analysis of Nanoscale Self-Assembled Discoidal Lipid Bilayers by Solid-State NMR Spectroscopy*, Biophys. J., **91**, 3819–3828, 2006.
45. I N Gorshkova, T Liu, H Y Kan, A Chroni, V I Zannis, and D Atkinson, *Structure and stability of apolipoprotein a-I in solution and in discoidal high-density lipoprotein probed by double charge ablation and deletion mutation*, Biochemistry, **45**, 1242–1254, 2006.
46. Amy Y. Shih, Anton Arkhipov, Peter L. Freddolino, Stephen G. Sligar, and Klaus Schulten, *Assembly of Lipids and Proteins into Lipoprotein Particles*, J. Phys. Chem. B, **111**, 11095–11104, 2007.
47. Amy Y. Shih, Peter L. Freddolino, Stephen G. Sligar, and Klaus Schulten, *Disassembly of Nanodiscs with Cholate*, Nano Lett., **7**, 1692–1696, 2007.

48. D I Svergun, C Barberato, and M H J Koch, *CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates*, J. Appl. Cryst., **28**, 768–773, 1995.
49. D. Sakamuro, K. J. Elliott, R. Wechsler-Reya, and G. C. Prendergast, *BIN1 is a novel MYC-interacting protein with features of a tumour suppressor*, Nat. Genet., **14**, 69–77, 1996.
50. G. Ren, P. Vajjhala, J. S. Lee, B. Winsor, and A. L. Munn, *The BAR Domain Proteins: Molding Membranes in Fission, Fusion, and Phagy*, Microbiology and molecular biology reviews, **70**, 37–120, 2006.
51. Brain J Peter, Helen M Kent, Ian G Mills, Yvonne Vallis, P. Johnathon G. Butler, Philip R. Evans, and Harvey T. McMahon, *BAR domains as sensors of membrane curvature: The amphiphysin BAR structure*, Science, **303**, 495–499, 2004.
52. P. D. Blood and G. A. Voth, *Direct observation of Bin/amphiphysin/Rvs (BAR) domain-induced membrane curvature by means of molecular dynamics simulations*, Proc. Natl. Acad. Sci. USA, **103**, 15068–15072, 2006.
53. James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten, *Scalable Molecular Dynamics with NAMD*, J. Comp. Chem., **26**, 1781–1802, 2005.
54. Ying Yin, Anton Arkhipov, Zhongzhou Chen, and Klaus Schulten, *Membrane tubulation by amphiphysin BAR domains*, 2007, Submitted.
55. Anton Arkhipov, Peter L. Freddolino, and Klaus Schulten, *Stability and Dynamics of Virus Capsids Described by Coarse-Grained Modeling*, Structure, **14**, 1767–1777, 2006.
56. Anton Arkhipov, Peter L. Freddolino, Katsumi Imada, Keiichi Namba, and Klaus Schulten, *Coarse-Grained Molecular Dynamics Simulations of a Rotating Bacterial Flagellum*, Biophys. J., **91**, 4589–4597, 2006.
57. William Humphrey, Andrew Dalke, and Klaus Schulten, *VMD – Visual Molecular Dynamics*, J. Mol. Graphics, **14**, 33–38, 1996.
58. Thomas Martinetz and Klaus Schulten, *Topology Representing Networks*, Neur. Netw., **7**, no. 3, 507–522, 1994.
59. A.D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. W. E. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, *All-atom empirical potential for molecular modeling and dynamics studies of proteins.*, J. Phys. Chem. B, **102**, 3586–3616, 1998.
60. Arnold J. Levine, *Viruses*, Scientific American Library, 1991.
61. S. J. Flint, L. W. Enquist, V. R. Racaniello, and A. M. Skalka, *Principles of Virology*, ASM Press, Washington, DC, 2 edition, 2004.
62. Peter L. Freddolino, Anton S. Arkhipov, Steven B. Larson, Alexander McPherson, and Klaus Schulten, *Molecular dynamics simulations of the complete satellite tobacco mosaic virus*, Structure, **14**, 437–449, 2006.
63. Robert W. Lucas, Steven B. Larson, and Alexander McPherson, *The crystallographic structure of brome mosaic virus*, J. Mol. Biol., **317**, 95–108, 2002.
64. Y. G. Kuznetsov, S. Daijogo, J. Zhou, B. L. Semler, and A. McPherson, *Atomic force*

microscopy analysis of icosahedral virus RNA, J. Mol. Biol., **347**, 41–52, 2005.
65. Howard C. Berg, *Motile behavior of bacteria*, Physics Today, **53**, 24–29, 2000.
66. Fadel A Samatey, Katsumi Imada, Shigehiro Nagashima, Ferenc Vonderviszt, Takashi Kumasaka, Masaki Yamamoto, and Keiichi Namba, *Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling*, Nature, **410**, 331–337, 2001.
67. Linda Turner, William S. Ryu, and Howard C. Berg, *Real-Time Imaging of Fluorescent Flagellar Filaments*, J. Bacteriol., **182**, no. 10, 2793–2801, 2000.
68. Jian Zhou, Ian F Thorpe, Sergey Izvekov, and Gregory A Voth, *Coarse-grained peptide modeling using a systematic multiscale approach.*, Biophys. J., **92**, no. 12, 4289–4303, Jun 2007.
69. Yeshitila Gebremichael, Gary S. Ayton, and Gregory A. Voth, *Mesoscopic Modeling of bacterial Flagellar Microhydrodynamics*, Biophys. J., **91**, 3640–3652, 2006.
70. Akio Kitao, Koji Yonekura, Saori Maki-Yonekura, Fadel A. Samatey, Katsumi Imada, Keiichi Namba, and Nobuhiro Go, *Switch interactions control energy frustration and multiple flagellar filament structures*, Proc. Natl. Acad. Sci. USA, **103**, no. 13, 4894–4899, 2006.
71. Brian N. Dominy and Charles L. Brooks, III, *Development of a generalized Born model parametrization for proteins and nucleic acids*, J. Phys. Chem. B, **103**, 3765–3773, 1999.
72. Donald Bashford and David A. Case, *Generalized Born Models of Macromolecular Solvation Effects*, Annu. Rev. Phys. Chem., **51**, 129–152, 2000.
73. John Mongan, David A. Case, and J. Andrew McCammon, *Constant pH molecular dynamics in generalized Born implicit solvent.*, J. Comp. Chem., **25**, 2038–2048, 2004.
74. B. J. Reynwar, G. Illya, V. A. Harmandaris, M. M. Müller, K. Kremer, and M. Deserno, *Aggregation and vesiculation of membrane proteins by curvature-mediated interactions*, Nature, **447**, 461–464, 2007.

Introduction to Multigrid Methods for Elliptic Boundary Value Problems

Arnold Reusken

Institut für Geometrie und Praktische Mathematik
RWTH Aachen, D-52056 Aachen, Germany
E-mail: reusken@igpm.rwth-aachen.de.

We treat multigrid methods for the efficient iterative solution of discretized elliptic boundary value problems. Two model problems are the Poisson equation and the Stokes problem. For the discretization we use standard finite element spaces. After discretization one obtains a large sparse linear system of equations. We explain multigrid methods for the solution of these linear systems. The basic concepts underlying multigrid solvers are discussed. Results of numerical experiments are presented which demonstrate the efficiency of these method. Theoretical convergence analyses are given that prove the typical grid independent convergence of multigrid methods.

1 Introduction

In these lecture notes we treat multigrid methods (MGM) for solving discrete elliptic boundary value problems. We assume that the reader is familiar with discretization methods for such partial differential equations. In our presentation we apply on finite element discretizations. We consider the following two model problems. Firstly, the Poisson equation

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega \subset \mathbb{R}^d, \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{1}$$

with f a (sufficiently smooth) source term and $d = 2$ or 3. The unknown is a scalar function u (for example, a temperature distribution) on Ω. We assume that the domain Ω is open, bounded and connected. The second problem consists of the Stokes equations

$$\begin{aligned} -\Delta \mathbf{u} + \nabla p &= \mathbf{f} \quad \text{in } \Omega \subset \mathbb{R}^d, \\ \operatorname{div} \mathbf{u} &= 0 \quad \text{in } \Omega, \\ \mathbf{u} &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{2}$$

The unknowns are the velocity vector function $\mathbf{u} = (u_1, \ldots, u_d)$ and the scalar pressure function p. To make this problem well-posed one needs an additional condition on p, for example, $\int_\Omega p\, dx = 0$. Both problems belong to the class of *elliptic boundary value problems*. Discretization of such partial differential equations using a finite difference, finite volume or finite element technique results in a *large sparse linear system of equations*. In the past three decades the development of *efficient iterative solvers* for such systems of equations has been an important research topic in numerical analysis and computational engineering. Nowadays it is recognized that multigrid iterative solvers are highly efficient for this type of problems and often have "optimal" complexity. There is an extensive literature on this subject. For a thorough treatment of multigrid methods we refer to the monograph

of Hackbusch[1]. For an introduction to multigrid methods requiring less knowledge of mathematics, we refer to Wesseling[2], Briggs[3], Trottenberg et al.[4]. A theoretical analysis of multigrid methods is presented in Bramble[5]. In these lecture notes we restrict ourselves to an introduction to the multigrid concept. We discuss several multigrid methods, heuristic concepts and theoretical analyses concerning convergence properties.

In the field of iterative solvers for discretized partial differential equations one can distinguish several classes of methods, namely *basic iterative methods* (eg., Jacobi, Gauss-Seidel), *Krylov subspace methods* (eg., CG, GMRES, BiCGSTAB) and *multigrid solvers*. For solving a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ which results from the discretization of an elliptic boundary value problem the first two classes need as input (only) the matrix $\mathbf{A}$ and the righthand side $\mathbf{b}$. The fact that these data correspond to a certain underlying continuous boundary value problem is *not* used in the iterative method. However, the relation between the data ($\mathbf{A}$ and $\mathbf{b}$) and the underlying problem can be useful for the development of a fast iterative solver. Due to the fact that $\mathbf{A}$ results from a discretization procedure we know, for example, that there are other matrices which, in a certain natural sense, are similar to the matrix $\mathbf{A}$. These matrices result from the discretization of the underlying continuous boundary value problem on other grids than the grid corresponding to the given discrete problem $\mathbf{A}\mathbf{x} = \mathbf{b}$. *The use of discretizations of the given continuous problem on several grids with different mesh sizes plays an important role in the multigrid concept*. Due to the fact that in multigrid methods discrete problems on different grids are needed, the implementation of multigrid methods is in general (much) more involved than the implementation of, for example, Krylov subspace methods. We also note that for multigrid methods it is relatively hard to develop "black box" solvers which are applicable to a wide class of problems. In recent years so-called *algebraic multigrid methods* have become quite popular. In these methods one tries to reduce the amount of geometric information (eg., different grids) that is needed in the solver, thus making the multigrid method more algebraic. We will not discuss such algebraic MGM in these lecture notes.

We briefly outline the contents. In section 2 we explain the main ideas of the MGM using a simple one dimensional problem. In section 3 we introduce multigrid methods for discretizations of *scalar* elliptic boundary value problems like the Poisson equation (1). In section 4 we present results of a numerical experiment with a standard multigrid solver applied to a discrete Poisson equation in 3D. In section 5 we introduce the main ideas for a multigrid method applied to a (generalized) Stokes problem. In section 6 we present results of a numerical experiments with a Stokes equation. In the final part of these notes, the sections 7 and 8, we present convergence analyses of these multigrid methods for the two classes of elliptic boundary value problems.

2 Multigrid for a One-Dimensional Model Problem

In this section we consider a simple model situation to show the basic principle behind the multigrid approach. We consider the two-point boundary value model problem

$$\begin{cases} -u''(x) = f(x), & x \in \Omega := (0,1). \\ u(0) = u(1) = 0\,. \end{cases} \tag{3}$$

We will use a finite element method for the discretization of this problem. This, however, is *not* essential: other discretization methods (finite differences, finite volumes) result in dis-

crete problems that are very similar. The corresponding multigrid methods have properties very similar to those in the case of a finite element discretization.

For the finite element discretization one needs a variational formulation of the boundary value problem in a suitable function space. We do not treat this issue here, but refer to the literature for information on this subject, eg. Hackbusch[6], Großmann[7]. For the two-point boundary value problem given above the appropriate function space is the Sobolov space $H_0^1(\Omega) := \{\, v \in L^2(\Omega) \mid v' \in L^2(\Omega),\ \ v(0) = v(1) = 0 \,\}$, where v' denotes a *weak* derivative of v. The variational formulation of the problem (3) is: find $u \in H_0^1(\Omega)$ such that

$$\int_0^1 u'v'\,dx = \int_0^1 fv\,dx \quad \text{for all } \ v \in H_0^1(\Omega).$$

For the discretization we introduce a sequence of nested uniform grids. For $\ell = 0, 1, 2, \ldots,$ we define

$$h_\ell = 2^{-\ell-1} \qquad \text{("mesh size")}\,, \tag{4}$$

$$n_\ell = h_\ell^{-1} - 1 \qquad \text{("number of interior grid points")}\,, \tag{5}$$

$$\xi_{\ell,i} = ih_\ell\,, \quad i = 0, 1, ..., n_\ell + 1 \qquad \text{("grid points")}\,, \tag{6}$$

$$\Omega_\ell^{\text{int}} = \{\xi_{\ell,i} \mid 1 \le i \le n_\ell\} \qquad \text{("interior grid")}\,, \tag{7}$$

$$\mathcal{T}_{h_\ell} = \cup\{\, [\xi_{\ell,i}, \xi_{\ell,i+1}] \mid 0 \le i \le n_\ell \,\} \qquad \text{("triangulation")}\,. \tag{8}$$

The space of *linear finite elements* corresponding to the triangulation $\mathcal{T}_{h_\ell}$ is given by

$$V_\ell := \{\, v \in C(\Omega) \mid v_{|[\xi_{\ell,i},\xi_{\ell,i+1}]} \in \mathcal{P}_1\,,\ i = 0, \ldots, n_\ell,\ \ v(0) = v(1) = 0 \,\}.$$

The standard nodal basis in this space is denoted by $(\phi_i)_{1\le i\le n_\ell}$. These functions satisfy $\phi_i(\xi_{\ell,i}) = 1,\ \phi_i(\xi_{\ell,j}) = 0$ for all $j \neq i$. This basis induces an isomorphism

$$P_\ell : \mathbb{R}^{n_\ell} \to V_\ell\,, \quad P_\ell \mathbf{x} = \sum_{i=1}^{n_\ell} x_i \phi_i. \tag{9}$$

The Galerkin discretization in the space V_ℓ is as follows: determine $u_\ell \in V_\ell$ such that

$$\int_0^1 u_\ell' v_\ell'\,dx = \int_0^1 fv_\ell\,dx \quad \text{for all } \ v_\ell \in V_\ell.$$

Using the representation $u_\ell = \sum_{j=1}^{n_\ell} x_j \phi_j$ this yields a linear system

$$\mathbf{A}_\ell \mathbf{x}_\ell = \mathbf{b}_\ell\,, \quad (A_\ell)_{ij} = \int_0^1 \phi_i' \phi_j'\,dx, \quad (b_\ell)_i = \int_0^1 f\phi_i\,dx. \tag{10}$$

The solution of this discrete problem is denoted by $\mathbf{x}_\ell^*$. The solution of the Galerkin discretization in the function space V_ℓ is given by $u_\ell = P_\ell \mathbf{x}_\ell^*$. A simple computation shows that

$$\mathbf{A}_\ell = h_\ell^{-1}\text{tridiag}(-1, 2, -1) \in \mathbb{R}^{n_\ell \times n_\ell}.$$

Note that, apart from a scaling factor, the same matrix results from a standard discretization with finite differences of the problem (3).

Clearly, in practice one should not solve the problem in (10) using an iterative method (a Cholesky factorization $\mathbf{A} = \mathbf{L}\mathbf{L}^T$ is stable and efficient). However, we do apply a basic

iterative method here, to illustrate a certain "smoothing" property which plays an important role in multigrid methods. We consider the damped Jacobi method

$$\mathbf{x}_\ell^{k+1} = \mathbf{x}_\ell^k - \frac{1}{2}\omega h_\ell(\mathbf{A}_\ell \mathbf{x}_\ell^k - \mathbf{b}_\ell) \quad \text{with} \quad \omega \in (0,1]\,. \tag{11}$$

The iteration matrix of this method, which describes the error propagation $\mathbf{e}_\ell^{k+1} = \mathbf{C}_\ell \mathbf{e}_\ell^k$, $\mathbf{e}_\ell^k := \mathbf{x}_\ell^* - \mathbf{x}_\ell^k$, is given by

$$\mathbf{C}_\ell = \mathbf{C}_\ell(\omega) = \mathbf{I} - \frac{1}{2}\omega h_\ell \mathbf{A}_\ell\,.$$

In this simple model problem an orthogonal eigenvector basis of $\mathbf{A}_\ell$, and thus of $\mathbf{C}_\ell$, too, is known. This basis is closely related to the "Fourier modes":

$$w^\nu(x) = \sin(\nu\pi x), \quad x \in [0,1], \quad \nu = 1, 2, \ldots\,.$$

Note that w^ν satisfies the boundary conditions in (3) and that $-(w^\nu)''(x) = (\nu\pi)^2 w^\nu(x)$ holds, and thus w^ν is an eigenfunction of the problem in (3). We introduce vectors $\mathbf{z}_\ell^\nu \in \mathbb{R}^{n_\ell}$, $1 \le \nu \le n_\ell$, which correspond to the Fourier modes w^ν restricted to the interior grid $\Omega_\ell^{\rm int}$:

$$\mathbf{z}_\ell^\nu := \left(w^\nu(\xi_{\ell,1}), w^\nu(\xi_{\ell,2}), \ldots, w^\nu(\xi_{\ell,n_\ell})\right)^T\,.$$

These vectors form an orthogonal basis of $\mathbb{R}^{n_\ell}$. For $\ell = 2$ we give an illustration in Fig. 1.

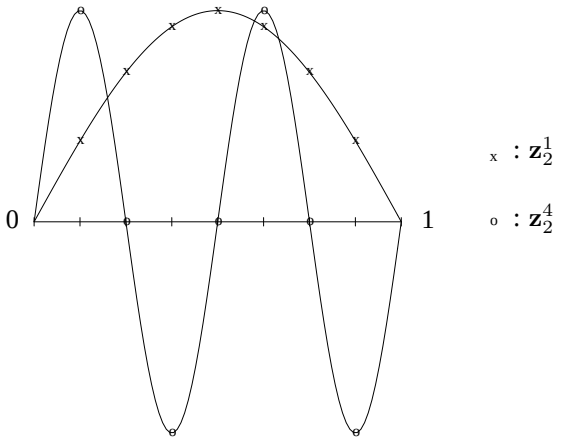


Figure 1. Two discrete Fourier modes.

To a vector $\mathbf{z}_\ell^\nu$ there corresponds a frequency ν. For $\nu < \frac{1}{2}n_\ell$ the vector $\mathbf{z}_\ell^\nu$, or the corresponding finite element function $P_\ell \mathbf{z}_\ell^\nu$, is called a "*low frequency mode*", and for $\nu \ge \frac{1}{2}n_\ell$ this vector [finite element function] is called a "*high frequency mode*". The vectors $\mathbf{z}_\ell^\nu$ are eigenvectors of the matrix $\mathbf{A}_\ell$:

$$\mathbf{A}_\ell \mathbf{z}_\ell^\nu = \frac{4}{h_\ell}\sin^2(\nu\frac{\pi}{2}h_\ell)\mathbf{z}_\ell^\nu,$$

and thus we have

$$\mathbf{C}_\ell \mathbf{z}_\ell^\nu = (1 - 2\omega \sin^2(\nu \frac{\pi}{2} h_\ell)) \mathbf{z}_\ell^\nu \ . \tag{12}$$

From this we obtain

$$\begin{aligned} \|\mathbf{C}_\ell\|_2 &= \max_{1 \le \nu \le n_\ell} \ |1 - 2\omega \sin^2(\nu \tfrac{\pi}{2} h_\ell)| \\ &= 1 - 2\omega \sin^2(\tfrac{\pi}{2} h_\ell) = 1 - \tfrac{1}{2}\omega \pi^2 h_\ell^2 + \mathcal{O}(h_\ell^4) \ . \end{aligned} \tag{13}$$

From this we see that the damped Jacobi method is convergent ($\|\mathbf{C}_\ell\|_2 < 1$), but that the rate of convergence will be very low for h_ℓ small.

Note that the eigenvalues and the eigenvectors of $\mathbf{C}_\ell$ are functions of $\nu h_\ell \in [0, 1]$:

$$\lambda_{\ell,\nu} := 1 - 2\omega \sin^2(\nu \frac{\pi}{2} h_\ell) =: g_\omega(\nu h_\ell) \ , \text{ with} \tag{14a}$$

$$g_\omega(y) \ = \ 1 - 2\omega \sin^2(\frac{\pi}{2} y), \qquad y \in [0, 1]. \tag{14b}$$

Hence, the size of the eigenvalues $\lambda_{\ell,\nu}$ can directly be obtained from the graph of the function g_ω. In Fig. 2 we show the graph of the function g_ω for a few values of ω.

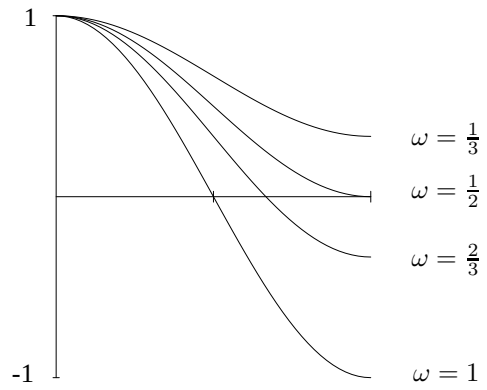


Figure 2. Graph of g_ω.

From the graphs in this figure we conclude that for a suitable choice of ω we have $|g_\omega(y)| \ll 1$ if $y \in [\frac{1}{2}, 1]$. We choose $\omega = \frac{2}{3}$ (then $|g_\omega(\frac{1}{2})| = |g_\omega(1)|$ holds). Then we have $|g_{\frac{2}{3}}(y)| \le \frac{1}{3}$ for $y \in [\frac{1}{2}, 1]$. Using this and the result in (14a) we obtain

$$|\lambda_{\ell,\nu}| \le \frac{1}{3} \quad \text{for } \nu \ge \frac{1}{2} n_\ell \ .$$

Hence:

the high frequency modes are strongly damped by the iteration matrix $\mathbf{C}_\ell$.

From Fig. 2 it is also clear that the low rate of convergence of the damped Jacobi method is caused by the low frequency modes ($\nu h_\ell \ll 1$).

Summarizing, we draw the conclusion that in this example the damped Jacobi method will "smooth" the error. This elementary observation is of great importance for the two-grid method introduced below. In the setting of multigrid methods the damped Jacobi method is called a "smoother". The smoothing property of damped Jacobi is illustrated in Fig. 3. It is important to note that the discussion above concerning smoothing is related to

Graph of a starting error.

Graph of the error after one damped Jacobi iteration ($\omega = \frac{2}{3}$).

Figure 3. Smoothing property of damped Jacobi.

the iteration matrix $\mathbf{C}_\ell$, which means that the *error* will be made smoother by the damped Jacobi method, but not (necessarily) the new iterant $\mathbf{x}^{k+1}$.

In multigrid methods we have to transform information from one grid to another. For that purpose we introduce so-called *prolongations* and *restrictions*. In a setting with nested finite element spaces these operators can be defined in a very natural way. Due to the nestedness the identity operator

$$I_\ell : V_{\ell-1}, \rightarrow V_\ell, \quad I_\ell v = v,$$

is well-defined. This identity operator represents linear interpolation as is illustrated for $\ell = 2$ in Fig. 4. The matrix representation of this interpolation operator is given by

$$\mathbf{p}_\ell : \mathbb{R}^{n_{\ell-1}} \rightarrow \mathbb{R}^{n_\ell}, \quad \mathbf{p}_\ell := P_\ell^{-1} P_{\ell-1}. \tag{15}$$

A simple computation yields

$$\mathbf{p}_\ell = \begin{bmatrix} \frac{1}{2} & & \emptyset \\ 1 & & \\ \frac{1}{2} & \frac{1}{2} & \\ & 1 & \\ & \frac{1}{2} & \\ & \ddots & \\ & & \frac{1}{2} \\ & & 1 \\ \emptyset & & \frac{1}{2} \end{bmatrix}_{n_\ell \times n_{\ell-1}} . \tag{16}$$

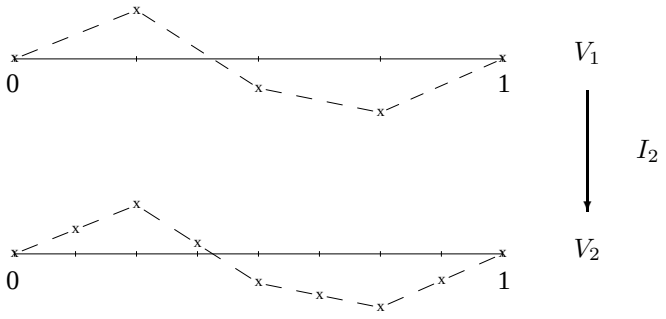


Figure 4. Canonical prolongation.

We can also restrict a given grid function v_ℓ on $\Omega_\ell^{\rm int}$ to a grid function on $\Omega_{\ell-1}^{\rm int}$. An obvious approach is to use a restriction r based on simple injection:

$$(r_{\rm inj} v_\ell)(\xi) = v_\ell(\xi) \quad \text{if} \ \ \xi \in \Omega_{\ell-1}^{\rm int} \, .$$

When used in a multigrid method then often this restriction based on injection is not satisfactory (cf. Hackbusch[1], section 3.5). A better method is obtained if a natural Galerkin property is satisfied. It can easily be verified (cf. also lemma 3.2) that with $\mathbf{A}_\ell, \ \mathbf{A}_{\ell-1}$ and $\mathbf{p}_\ell$ as defined in (10), (15) we have

$$\mathbf{r}_\ell \mathbf{A}_\ell \mathbf{p}_\ell = \mathbf{A}_{\ell-1} \quad \text{iff} \quad \mathbf{r}_\ell = \mathbf{p}_\ell^T . \tag{17}$$

Thus the natural Galerkin condition $\mathbf{r}_\ell \mathbf{A}_\ell \mathbf{p}_\ell = \mathbf{A}_{\ell-1}$ implies the choice

$$\mathbf{r}_\ell = \mathbf{p}_\ell^T \tag{18}$$

for the restriction operator.

The *two-grid* method is based on the idea that a smooth error, which results from the application of one or a few damped Jacobi iterations, can be approximated fairly well on a *coarser* grid. We now introduce this two-grid method.

Consider $\mathbf{A}_\ell \mathbf{x}_\ell^* = \mathbf{b}_\ell$ and let $\overline{\mathbf{x}}_\ell$ be the result of one or a few damped Jacobi iterations applied to a given starting vector $\mathbf{x}_\ell^0$. For the error $\mathbf{e}_\ell := \mathbf{x}_\ell^* - \overline{\mathbf{x}}_\ell$ we have

$$\mathbf{A}_\ell \mathbf{e}_\ell = \mathbf{b}_\ell - \mathbf{A}_\ell \overline{\mathbf{x}}_\ell =: \mathbf{d}_\ell \quad (\text{ “residual” or “defect”}). \tag{19}$$

Based on the assumption that $\mathbf{e}_\ell$ is smooth it seems reasonable to make the approximation $\mathbf{e}_\ell \approx \mathbf{p}_\ell \tilde{\mathbf{e}}_{\ell-1}$ with an appropriate vector (grid function) $\tilde{\mathbf{e}}_{\ell-1} \in \mathbb{R}^{n_{\ell-1}}$. To determine the vector $\tilde{\mathbf{e}}_{\ell-1}$ we use the equation (19) and the Galerkin property (17). This results in the equation

$$\mathbf{A}_{\ell-1} \tilde{\mathbf{e}}_{\ell-1} = \mathbf{r}_\ell \mathbf{d}_\ell$$

for the vector $\tilde{\mathbf{e}}_{\ell-1}$. Note that $\mathbf{x}^* = \overline{\mathbf{x}}_\ell + \mathbf{e}_\ell \approx \overline{\mathbf{x}}_\ell + \mathbf{p}_\ell \tilde{\mathbf{e}}_{\ell-1}$. Thus for the new iterant we take $\mathbf{x}_\ell := \overline{\mathbf{x}}_\ell + \mathbf{p}_\ell \tilde{\mathbf{e}}_{\ell-1}$. In a more compact formulation this two-grid method is as

follows:

$$
\left\{
\begin{array}{l}
\mathit{procedure}\ \mathrm{TGM}_\ell(\mathbf{x}_\ell, \mathbf{b}_\ell) \\
\quad \text{if } \ell = 0 \ \text{ then } \ \mathbf{x}_0 := \mathbf{A}_0^{-1}\mathbf{b}_0 \ \text{ else} \\
\quad \text{begin} \\
\qquad \mathbf{x}_\ell := J_\ell^\nu(\mathbf{x}_\ell, \mathbf{b}_\ell) \ \ (* \ \nu \text{ smoothing it., e.g. damped Jacobi } *) \\
\qquad \mathbf{d}_{\ell-1} := \mathbf{r}_\ell(\mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}_\ell) \quad (* \text{ restriction of defect } *) \\
\qquad \tilde{\mathbf{e}}_{\ell-1} := \mathbf{A}_{\ell-1}^{-1}\mathbf{d}_{\ell-1} \quad (* \text{ solve coarse grid problem } *) \\
\qquad \mathbf{x}_\ell := \mathbf{x}_\ell + \mathbf{p}_\ell \tilde{\mathbf{e}}_{\ell-1} \quad (* \text{ add correction } *) \\
\qquad \mathrm{TGM}_\ell := \mathbf{x}_\ell \\
\quad \text{end;}
\end{array}
\right. \tag{20}
$$

Often, after the coarse grid correction $\mathbf{x}_\ell := \mathbf{x}_\ell + \mathbf{p}_\ell \tilde{\mathbf{e}}_{\ell-1}$, one or a few smoothing iterations are applied. Smoothing before/after the coarse grid correction is called pre/post-smoothing. Besides the smoothing property a second property which is of great importance for a multigrid method is the following:

The coarse grid system $\mathbf{A}_{\ell-1}\tilde{\mathbf{e}}_{\ell-1} = \mathbf{d}_{\ell-1}$ is of the same form as the system $\mathbf{A}_\ell \mathbf{x}_\ell = \mathbf{b}_\ell$.

Thus for solving the problem $\mathbf{A}_{\ell-1}\tilde{\mathbf{e}}_{\ell-1} = \mathbf{d}_{\ell-1}$ *approximately* we can apply the two-grid algorithm in (20) recursively. This results in the following *multigrid method* for solving $\mathbf{A}_\ell \mathbf{x}_\ell^* = \mathbf{b}_\ell$:

$$
\left\{
\begin{array}{l}
\mathit{procedure}\ \mathrm{MGM}_\ell(\mathbf{x}_\ell, \mathbf{b}_\ell) \\
\quad \text{if } \ \ell = 0 \ \text{ then } \ \mathbf{x}_0 := \mathbf{A}_0^{-1}\mathbf{b}_0 \ \text{ else} \\
\quad \text{begin} \\
\qquad \mathbf{x}_\ell := J_\ell^{\nu_1}(\mathbf{x}_\ell, \mathbf{b}_\ell) \quad (* \text{ presmoothing } *) \\
\qquad \mathbf{d}_{\ell-1} := \mathbf{r}_\ell(\mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}_\ell) \\
\qquad \mathbf{e}_{\ell-1}^0 := \mathbf{0}; \ \text{ for } i = 1 \text{ to } \tau \text{ do } \ \mathbf{e}_{\ell-1}^i := \mathrm{MGM}_{\ell-1}(\mathbf{e}_{\ell-1}^{i-1}, \mathbf{d}_{\ell-1}); \\
\qquad \mathbf{x}_\ell := \mathbf{x}_\ell + \mathbf{p}_\ell \mathbf{e}_{\ell-1}^\tau \\
\qquad \mathbf{x}_\ell := J_\ell^{\nu_2}(\mathbf{x}_\ell, \mathbf{b}_\ell) \quad (* \text{ postsmoothing } *) \\
\qquad \mathrm{MGM}_\ell := \mathbf{x}_\ell \\
\quad \text{end;}
\end{array}
\right. \tag{21}
$$

If one wants to solve the system on a given finest grid, say with level number $\overline{\ell}$, i.e. $\mathbf{A}_{\overline{\ell}}\mathbf{x}_{\overline{\ell}}^* = \mathbf{b}_{\overline{\ell}}$, then we apply some iterations of $\mathrm{MGM}_{\overline{\ell}}(\mathbf{x}_{\overline{\ell}}, \mathbf{b}_{\overline{\ell}})$.

Based on efficiency considerations (cf. section 3) we usually take $\tau = 1$ ("V-cycle") or $\tau = 2$ ("W-cycle") in the recursive call in (21). For the case $\ell = 3$ the structure of one multigrid iteration with $\tau \in \{1, 2\}$ is illustrated in Fig. 5.

3 Multigrid for Scalar Elliptic Problems

In this section we introduce multigrid methods which can be used for solving discretized scalar elliptic boundary value problems. A model example from this problem class is the Poisson equation in (1). Opposite to the CG method, the applicability of multigrid methods

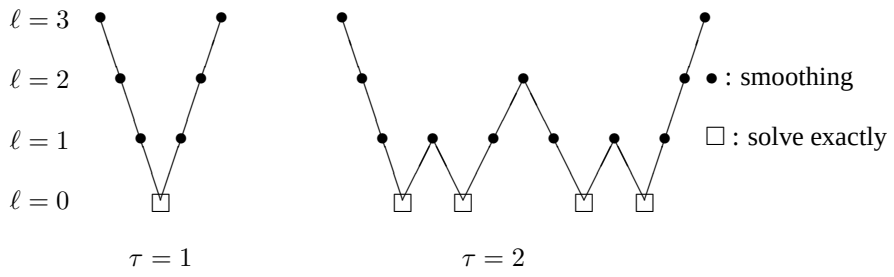


Figure 5. Structure of one multigrid iteration

is not restricted to symmetric problems. Multigrid methods can also be used for solving problems which are nonsymmetric (i.e., convection terms are present in the equation). If the problem is convection-dominated (the corresponding stiffness matrix then is strongly nonsymmetric) one usually has to modify the standard multigrid approach in the sense that special smoothers and/or special prolongations and restrictions should be used. We do not discuss this issue here.

We will introduce the two-grid and multigrid method by generalizing the approach of section 2 to the higher (i.e., two and three) dimensional case. We consider a scalar elliptic boundary value problems of the form

$$\begin{aligned} -\nabla \cdot (a\nabla u) + \mathbf{b} \cdot \nabla u + cu &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

This problem is considered in a domain $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3. We assume that the functions a, c and the vector function $\mathbf{b}$ are sufficiently smooth on Ω and

$$a(x) \geq a_0 > 0, \;\; c(x) - \frac{1}{2}\mathrm{div}\,\mathbf{b}(x) \geq 0 \quad \text{for all } \; x \in \bar{\Omega}. \tag{22}$$

These assumptions guarantee that the problem is elliptic and well-posed. In view of the finite element discretization we introduce the variational formulation of this problem. For this we need the Sobolov space $H_0^1(\Omega) := \{\, v \in L^2(\Omega) \mid \frac{\partial v}{\partial x_i} \in L^2(\Omega),\; i = 1, \ldots, d, \;\; v_{|\partial\Omega} = 0 \,\}$. The partial derivative $\frac{\partial v}{\partial x_i}$ has to be interpreted in a suitable weak sense. The variational formulation is as follows:

$$\begin{cases} \text{find } \; u \in H_0^1(\Omega) \;\; \text{such that} \\ k(u, v) = f(v) \quad \text{for all } \; v \in H_0^1(\Omega), \end{cases} \tag{23}$$

with a bilinear form and righthand side

$$k(u, v) = \int_\Omega a\nabla u^T \nabla v \, + \, \mathbf{b} \cdot \nabla u v \, + \, cuv \, dx. \, , \quad f(v) = \int_\Omega fv \, dx.$$

If (22) holds then this bilinear form is *continuous and elliptic* on $H_0^1(\Omega)$, i.e. there exist constants $\gamma > 0$ and c such that

$$k(u, u) \geq \gamma |u|_1^2, \quad k(u, v) \leq c|u|_1 |v|_1 \quad \text{for all } \; u, v \in H_0^1(\Omega).$$

Here we use $|u|_1 := \left(\int_\Omega \nabla u^T \nabla u \, dx \right)^{\frac{1}{2}}$, which is a norm on $H_0^1(\Omega)$. For the discretization of this problem we use simplicial finite elements. Let $\{\mathcal{T}_h\}$ be a regular family of triangulations of Ω consisting of d-simplices and V_h a corresponding finite element space. For simplicity we only consider *linear* finite elements:

$$V_h = \{ v \in C(\Omega) \mid v_{|T} \in \mathcal{P}_1 \ \text{ for all } \ T \in \mathcal{T}_h \}.$$

The presentation and implementation of the multigrid method is greatly simplified if we assume a given sequence of *nested* finite element spaces.

Assumption 3.1 *In the remainder we always assume that we have a sequence V_ℓ, $\ell = 0, 1, \ldots$, of simplicial finite element spaces which are nested:*

$$V_\ell \subset V_{\ell+1} \quad \textit{for all } \ \ell. \tag{24}$$

We note that this assumption is not necessary for a successful application of multigrid methods. For a treatment of multigrid methods in case of non-nestedness we refer to Trottenberg et al.[4]. The construction of a hierarchy of triangulations such that the corresponding finite element spaces are nested is discussed in Bey[8].

In V_ℓ we use the standard nodal basis $(\phi_i)_{1 \le i \le n_\ell}$. This basis induces an isomorphism

$$P_\ell : \mathbb{R}^{n_\ell} \to V_\ell \ , \quad P_\ell \mathbf{x} = \sum_{i=1}^{n_\ell} x_i \phi_i.$$

The Galerkin discretization: Find $u_\ell \in V_\ell$ such that

$$k(u_\ell, v_\ell) = f(v_\ell) \quad \text{for all } \ v_\ell \in V_\ell \tag{25}$$

can be represented as a linear system

$$\mathbf{A}_\ell \mathbf{x}_\ell = \mathbf{b}_\ell \, , \ \text{with } \ (A_\ell)_{ij} = k(\phi_j, \phi_i), \ (b_\ell)_i = f(\phi_i), \ \ 1 \le i, j \le n_\ell. \tag{26}$$

The solution $\mathbf{x}_\ell^*$ of this linear system yields the Galerkin finite element solution $u_\ell = P_\ell \mathbf{x}_\ell^*$. Along the same lines as in the one-dimensional case we introduce a multigrid method for solving this system of equations on an arbitrary level $\ell \ge 0$.

For the *smoother* we use a basic iterative method such as, for example, a *Richardson method*

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \omega_\ell (\mathbf{A}_\ell \mathbf{x}^k - \mathbf{b}),$$

a *damped Jacobi method*

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \omega \mathbf{D}_\ell^{-1} (\mathbf{A}_\ell \mathbf{x}^k - \mathbf{b}), \tag{27}$$

or a *Gauss-Seidel method*

$$\mathbf{x}^{k+1} = \mathbf{x}^k - (\mathbf{D}_\ell - \mathbf{L}_\ell)^{-1} (\mathbf{A}_\ell \mathbf{x}^k - \mathbf{b}), \tag{28}$$

where $\mathbf{D}_\ell - \mathbf{L}_\ell$ is the lower triangular part of the matrix $\mathbf{A}_\ell$. For such a method we use the general notation

$$\mathbf{x}^{k+1} = \mathcal{S}_\ell(\mathbf{x}^k, \mathbf{b}_\ell) = \mathbf{x}^k - \mathbf{M}_\ell^{-1} (\mathbf{A}_\ell \mathbf{x}^k - \mathbf{b}) \ , \quad k = 0, 1, \ldots$$

The corresponding iteration matrix is denoted by

$$\mathbf{S}_\ell = \mathbf{I} - \mathbf{M}_\ell^{-1} \mathbf{A}_\ell.$$

For the *prolongation* we use the matrix representation of the identity $I_\ell : V_{\ell-1} \to V_\ell$, i.e.,

$$\mathbf{p}_\ell := P_\ell^{-1} P_{\ell-1}. \tag{29}$$

The choice of the restriction is based on the following elementary lemma:

Lemma 3.2 *Let $\mathbf{A}_\ell$, $\ell \geq 0$, be the stiffness matrix defined in* (26) *and $\mathbf{p}_\ell$ as in* (29)*. Then for $\mathbf{r}_\ell : \mathbb{R}^{n_\ell} \to \mathbb{R}^{n_{\ell-1}}$ we have:*

$$\mathbf{r}_\ell \mathbf{A}_\ell \mathbf{p}_\ell = \mathbf{A}_{\ell-1} \quad \textit{if and only if} \quad \mathbf{r}_\ell = \mathbf{p}_\ell^T.$$

Proof: For the stiffness matrix matrix the identity

$$\langle \mathbf{A}_\ell \mathbf{x}, \mathbf{y} \rangle = k(P_\ell \mathbf{x}, P_\ell \mathbf{y}) \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^{n_\ell}$$

holds. From this we get

$$\begin{aligned} & \mathbf{r}_\ell \mathbf{A}_\ell \mathbf{p}_\ell = \mathbf{A}_{\ell-1} \\ \Leftrightarrow \quad & \langle \mathbf{A}_\ell \mathbf{p}_\ell \mathbf{x}, \mathbf{r}_\ell^T \mathbf{y} \rangle = \langle \mathbf{A}_{\ell-1} \mathbf{x}, \mathbf{y} \rangle \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^{n_{\ell-1}} \\ \Leftrightarrow \quad & k(P_{\ell-1}\mathbf{x}, P_\ell \mathbf{r}_\ell^T \mathbf{y}) = k(P_{\ell-1}\mathbf{x}, P_{\ell-1}\mathbf{y}) \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^{n_{\ell-1}}. \end{aligned}$$

Using the ellipticity of $k(\cdot, \cdot)$ it now follows that

$$\begin{aligned} & \mathbf{r}_\ell \mathbf{A}_\ell \mathbf{p}_\ell = \mathbf{A}_{\ell-1} \\ \Leftrightarrow \quad & P_\ell \mathbf{r}_\ell^T \mathbf{y} = P_{\ell-1}\mathbf{y} \quad \text{for all } \mathbf{y} \in \mathbb{R}^{n_{\ell-1}} \\ \Leftrightarrow \quad & \mathbf{r}_\ell^T \mathbf{y} = P_\ell^{-1} P_{\ell-1}\mathbf{y} = \mathbf{p}_\ell \mathbf{y} \quad \text{for all } \mathbf{y} \in \mathbb{R}^{n_{\ell-1}} \\ \Leftrightarrow \quad & \mathbf{r}_\ell^T = \mathbf{p}_\ell. \end{aligned}$$

Thus the claim is proved. ∎

This motivates that for the *restriction* we take:

$$\mathbf{r}_\ell := \mathbf{p}_\ell^T. \tag{30}$$

Using these components we can define a multigrid method with exactly the same structure as in (21):

```
procedure MGM_ℓ(x_ℓ, b_ℓ)
  if ℓ = 0 then x_0 := A_0^{-1} b_0 else
  begin
    x_ℓ := S_ℓ^{ν1}(x_ℓ, b_ℓ)   (* presmoothing *)
    d_{ℓ-1} := r_ℓ(b_ℓ − A_ℓ x_ℓ)
    e^0_{ℓ-1} := 0; for i = 1 to τ do e^i_{ℓ-1} := MGM_{ℓ-1}(e^{i-1}_{ℓ-1}, d_{ℓ-1});
    x_ℓ := x_ℓ + p_ℓ e^τ_{ℓ-1}
    x_ℓ := S_ℓ^{ν2}(x_ℓ, b_ℓ)   (* postsmoothing *)
    MGM_ℓ := x_ℓ
  end;
```

(31)

We briefly comment on some important issues related to this multigrid method.

Smoothers
For many problems basic iterative methods provide good smoothers. In particular the Gauss-Seidel method is often a very effective smoother. Other smoothers used in practice are the damped Jacobi method and the ILU method.

Prolongation and restriction
If instead of a discretization with nested finite element spaces one uses a finite difference or a finite volume method then one can not use the approach in (29) to define a prolongation. However, for these cases other canonical constructions for the prolongation operator exist. We refer to Hackbusch[1], Trottenberg et al.[4] or Wesseling[2] for a treatment of this topic. A general technique for the construction of a prolongation operator in case of nonnested finite element spaces is given in Braess[9].

Arithmetic costs per iteration
We discuss the arithmetic costs of one MGM_ℓ iteration as defined in (31). For this we introduce a unit of arithmetic work on level ℓ:

$$WU_\ell := \#\text{ flops needed for } \mathbf{A}_\ell \mathbf{x}_\ell - \mathbf{b}_\ell \text{ computation.} \tag{32}$$

We assume:

$$WU_{\ell-1} \lesssim g\, WU_\ell \quad \text{with } g < 1 \text{ independent of } \ell. \tag{33}$$

Note that if $\mathcal{T}_\ell$ is constructed through a uniform global grid refinement of $\mathcal{T}_{\ell-1}$ (for $n = 2$: subdivision of each triangle $T \in \mathcal{T}_{\ell-1}$ into four smaller triangles by connecting the midpoints of the edges) then (33) holds with $g = (\frac{1}{2})^d$. Furthermore we make the following assumptions concerning the arithmetic costs of each of the substeps in the procedure MGM_ℓ:

$$\left.\begin{aligned} \mathbf{x}_\ell &:= \mathcal{S}_\ell(\mathbf{x}_\ell, \mathbf{b}_\ell) \;:\; \text{costs} \lesssim WU_\ell \\ \mathbf{d}_{\ell-1} &:= \mathbf{r}_\ell(\mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}_\ell) \\ \mathbf{x}_\ell &:= \mathbf{x}_\ell + \mathbf{p}_\ell \mathbf{e}_{\ell-1}^\tau \end{aligned}\right\} \text{total costs} \lesssim 2\, WU_\ell$$

For the amount of work in one multigrid V-cycle ($\tau = 1$) on level ℓ, which is denoted by VMG_ℓ, we get using $\nu := \nu_1 + \nu_2$:

$$\begin{aligned} VMG_\ell &\lesssim \nu WU_\ell + 2WU_\ell + VMG_{\ell-1} = (\nu+2)WU_\ell + VMG_{\ell-1} \\ &\lesssim (\nu+2)\big(WU_\ell + WU_{\ell-1} + \ldots + WU_1\big) + VMG_0 \\ &\lesssim (\nu+2)\big(1 + g + \ldots + g^{\ell-1}\big)WU_\ell + VMG_0 \\ &\lesssim \frac{\nu+2}{1-g}\, WU_\ell. \end{aligned} \tag{34}$$

In the last inequality we assumed that the costs for computing $\mathbf{x}_0 = \mathbf{A}_0^{-1}\mathbf{b}_0$ (i.e., VMG_0) are negligible compared to WU_ℓ. The result in (34) shows that the arithmetic costs for one V-cycle are proportional (if $\ell \to \infty$) to the costs of a residual computation. For example, for $g = \frac{1}{8}$ (uniform refinement in 3D) the arithmetic costs of a V-cycle with $\nu_1 = \nu_2 = 1$ on level ℓ are comparable to $4\frac{1}{2}$ times the costs of a residual computation on level ℓ.

For the W-cycle ($\tau = 2$) the arithmetic costs on level ℓ are denoted by WMG_ℓ. We have:

$$\begin{aligned} WMG_\ell &\lesssim \nu WU_\ell + 2WU_\ell + 2WMG_{\ell-1} = (\nu+2)WU_\ell + 2WMG_{\ell-1} \\ &\lesssim (\nu+2)\big(WU_\ell + 2WU_{\ell-1} + 2^2 WU_{\ell-2} + \ldots + 2^{\ell-1}WU_1\big) + WMG_0 \\ &\lesssim (\nu+2)\big(1 + 2g + (2g)^2 + \ldots + (2g)^{\ell-1}\big)WU_\ell + WMG_0. \end{aligned}$$

From this we see that to obtain a bound proportional to WU_ℓ we have to assume

$$g < \frac{1}{2}.$$

Under this assumption we get for the W-cycle

$$WMG_\ell \lesssim \frac{\nu+2}{1-2g}\, WU_\ell$$

(again we neglected WMG_0). Similar bounds can be obtained for $\tau \geq 3$, provided $\tau g < 1$ holds.

3.1 Nested Iteration

We consider a sequence of discretizations of a given boundary value problem, as for example in (26):

$$\mathbf{A}_\ell \mathbf{x}_\ell = \mathbf{b}_\ell, \qquad \ell = 0, 1, 2, \ldots.$$

We assume that for a certain $\ell = \overline{\ell}$ we want to compute the solution $\mathbf{x}_{\overline{\ell}}^*$ of the problem $\mathbf{A}_{\overline{\ell}}\mathbf{x}_{\overline{\ell}} = \mathbf{b}_{\overline{\ell}}$ using an iterative method (not necessarily a multigrid method). In the nested iteration method we use the systems on coarse grids to obtain a *good starting vector* $\mathbf{x}_{\overline{\ell}}^0$ for this iterative method with relatively low computational costs. The nested iteration method for the computation of this starting vector $\mathbf{x}_{\overline{\ell}}^0$ is as follows

$$\left\{ \begin{array}{l} \text{compute the solution } \mathbf{x}_0^* \text{ of } \mathbf{A}_0\mathbf{x}_0 = \mathbf{b}_0 \\ \mathbf{x}_1^0 := \tilde{\mathbf{p}}_1 \mathbf{x}_0^* \quad \text{(prolongation of } \mathbf{x}_0^*\text{)} \\ \mathbf{x}_1^k := \text{result of } k \text{ iterations of an iterative method} \\ \qquad \text{applied to } \mathbf{A}_1\mathbf{x}_1 = \mathbf{b}_1 \text{ with starting vector } \mathbf{x}_1^0 \\ \mathbf{x}_2^0 := \tilde{\mathbf{p}}_2 \mathbf{x}_1^k \quad \text{(prolongation of } \mathbf{x}_1^k\text{)} \\ \mathbf{x}_2^k := \text{ result of } k \text{ iterations of an iterative method} \\ \qquad \text{applied to } \mathbf{A}_2\mathbf{x}_2 = \mathbf{b}_2 \text{ with starting vector } \mathbf{x}_2^0 \\ \vdots \\ \text{etc.} \\ \vdots \\ \mathbf{x}_{\overline{\ell}}^0 := \tilde{\mathbf{p}}_{\overline{\ell}} \mathbf{x}_{\overline{\ell}-1}^k \,. \end{array} \right. \tag{35}$$

In this nested iteration method we use a prolongation $\tilde{\mathbf{p}}_\ell : \mathbb{R}^{n_{\ell-1}} \to \mathbb{R}^{n_\ell}$. The nested iteration principle is based on the idea that $\tilde{\mathbf{p}}_\ell \mathbf{x}_{\ell-1}^*$ is expected to be a reasonable approximation of $\mathbf{x}_\ell^*$, because $\mathbf{A}_{\ell-1}\mathbf{x}_{\ell-1}^* = \mathbf{b}_{\ell-1}$ and $\mathbf{A}_\ell \mathbf{x}_\ell^* = \mathbf{b}_\ell$ are discretizations of the same

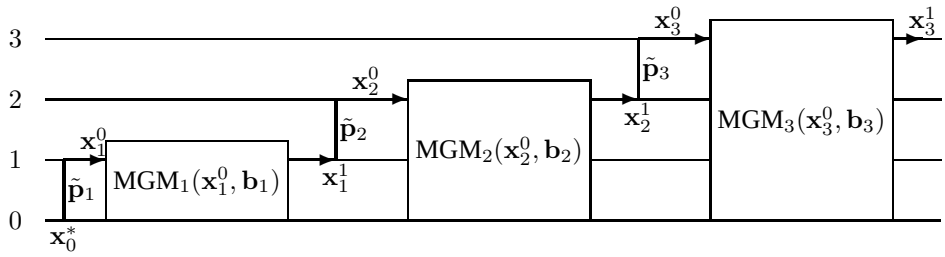


Figure 6. Multigrid and nested iteration.

continuous problem. With respect to the computational costs of this approach we note the following (cf. Hackbusch[1], section 5.3). For the nested iteration to be a feasible approach, the number of iterations applied on the coarse grids (i.e. k in (35)) should not be "too large" and the number of grid points in the union of all coarse grids (i.e. level $0, 1, 2, ..., \overline{\ell} - 1$) should be at most of the same order of magnitude as the number of grid points in the level $\overline{\ell}$ grid. Often, if one uses a multigrid solver these two conditions are satisfied. Usually in multigrid we use coarse grids such that the number of grid points decreases in a geometric fashion, and for k in (35) we can often take $k = 1$ or $k = 2$ due to the fact that on the coarse grids we use the multigrid method, which has a high rate of convergence.

If one uses the algorithm $\mathrm{MGM}_{\overline{\ell}}$ from (31) as the solver on level $\overline{\ell}$ then the implementation of the nested iteration method can be realized with only little additional effort because the coarse grid data structure and coarse grid operators (e.g. $\mathbf{A}_\ell,\ \ell < \overline{\ell}$) needed in the nested iteration method are already available.

If in the nested iteration method we use a multigrid iterative solver on all levels we obtain the following algorithmic structure:

$$\left\{\begin{array}{l} \mathbf{x}_0^* := \mathbf{A}_0^{-1}\mathbf{b}_0;\ \mathbf{x}_0^k := \mathbf{x}_0^* \\ \text{for}\ \ \ell = 1 \text{ to } \overline{\ell} \text{ do} \\ \quad \text{begin} \\ \quad\quad \mathbf{x}_\ell^0 := \tilde{\mathbf{p}}_\ell \mathbf{x}_{\ell-1}^k \\ \quad\quad \text{for}\ \ i = 1 \text{ to } k \text{ do}\ \ \mathbf{x}_\ell^i := \mathrm{MGM}_\ell(\mathbf{x}_\ell^{i-1}, \mathbf{b}_\ell) \\ \quad \text{end;} \end{array}\right. \tag{36}$$

For the case $\overline{\ell} = 3$ and $k = 1$ this method is illustrated in Fig. 6.

Remark 3.3 The prolongation $\tilde{\mathbf{p}}_\ell$ used in the nested iteration may be the same as the prolongation $\mathbf{p}_\ell$ used in the multigrid method. However, from the point of view of efficiency it is sometimes better to use in the nested iteration a prolongation $\tilde{\mathbf{p}}_\ell$ that has a higher order of accuracy than the prolongation used in the multigrid method. □

4 Numerical Experiment: Multigrid Applied to a Poisson Equation

In this section we present results of a standard multigrid solver applied to the model problem of the Poisson equation:

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega := (0,1)^3, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

We take $f(x_1, x_2, x_3) = x_1^2 + e^{x_2} x_1 + x_3^2 x_2$. For the discretization we start with a uniform subdivision of Ω into cubes with edges of length $h_0 := \frac{1}{4}$. Each cube is subdivided into six tetrahedra. This yields the starting triangulation $\mathcal{T}_0$ of Ω. The triangulation $\mathcal{T}_1$ with mesh size $h_1 = \frac{1}{8}$ is constructed by regular subdivision of each tetrahedron in $\mathcal{T}_0$ into 8 child tetrahedra. This uniform refinement strategy is repeated, resulting in a family of triangulations $(\mathcal{T}_\ell)_{\ell \geq 0}$ with corresponding mesh size $h_\ell = 2^{-\ell-2}$. For discretization of this problem we use the space of linear finite elements on these triangulations. The resulting linear system is denoted by $\mathbf{A}_\ell \mathbf{x}_\ell = \mathbf{b}_\ell$. We consider the problem of solving this linear system on a fixed finest level $\ell = \bar{\ell}$. Below we consider $\bar{\ell} = 1, \ldots, 5$. For $\bar{\ell} = 5$ the triangulation contains 14.380.416 tetrahedra and in the linear system we have 2.048.383 unknowns.

We briefly discuss the components used in the multigrid method for solving this linear system. For the prolongation and restriction we use the canonical ones as in (29), (30). For the smoother we use two different methods, namely a damped Jacobi method and a symmetric Gauss-Seidel method (SGS). The damped Jacobi method is as in (27) with $\omega := 0.7$. The symmetric Gauss-Seidel method consists of two substeps. In the first step we use a Gauss-Seidel iteration as in (28). In the second step we apply this method with a reversed ordering of the equations and the unknowns. The arithmetic costs per iteration for such a symmetric Gauss-Seidel smoother are roughly twice as high as for a damped Jacobi method. In the experiment we use the same number of pre- and post-smoothing iterations, i.e. $\nu_1 = \nu_2$. The total number of smoothing iterations per multigrid iteration is $\nu := \nu_1 + \nu_2$. We use a multigrid V-cycle. i.e., $\tau = 1$ in the recursive call in (31). The coarsest grid used in the multigrid method is $\mathcal{T}_0$, i.e. with a mesh size $h_0 = \frac{1}{4}$. In all experiments we use a starting vector $\mathbf{x}^0 := 0$. The rate of convergence is measured by looking at relative residuals:

$$r_k := \frac{\|\mathbf{A}_{\bar{\ell}} \mathbf{x}^k - \mathbf{b}_{\bar{\ell}}\|_2}{\|\mathbf{b}_{\bar{\ell}}\|_2}.$$

In Fig. 7 (left) we show results for SGS with $\nu = 4$. For $\bar{\ell} = 1, \ldots, 5$ we plotted the relative residuals r_k for $k = 1, \ldots, 8$. In Fig. 7 (right) we show results for the SGS method with varying number of smoothing iterations, namely $\nu = 2, 4, 6$. For $\bar{\ell} = 1, \ldots, 5$ we give the average residual reduction per iteration $r := (r_8)^{\frac{1}{8}}$.

These results show the very fast and essentially level independent rate of convergence of this multigrid method. For a larger number of smoothing iterations the convergence is faster. On the other hand, also the costs per iteration then increase, cf. (34) (with $g = \frac{1}{8}$). Usually, in practice the number of smoothings per iteration is not taken very large. Typical values are $\nu = 2$ or $\nu = 4$. In the Fig. 8 we show similar results but now for the damped Jacobi smoother (damping with $\omega = 0.7$) instead of the SGS method.

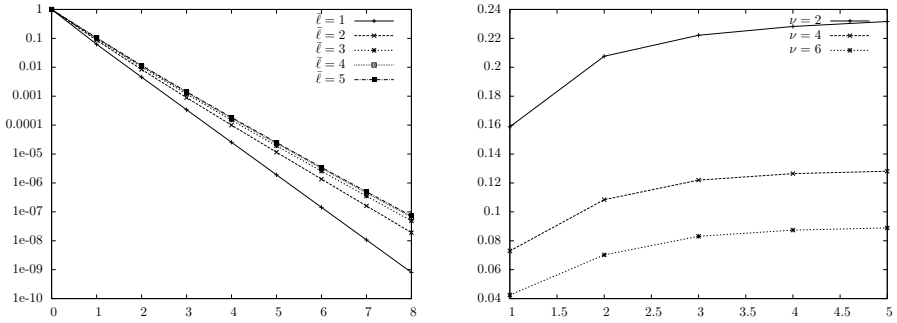


Figure 7. Convergence of multigrid V-cycle with SGS smoother. Left: r_k, for $k = 0, \ldots, 8$ and $\bar{\ell} = 1, \ldots, 5$. Right: $(r_8)^{\frac{1}{8}}$ for $\bar{\ell} = 1, \ldots, 5$ and $\nu = 2, 4, 6$.

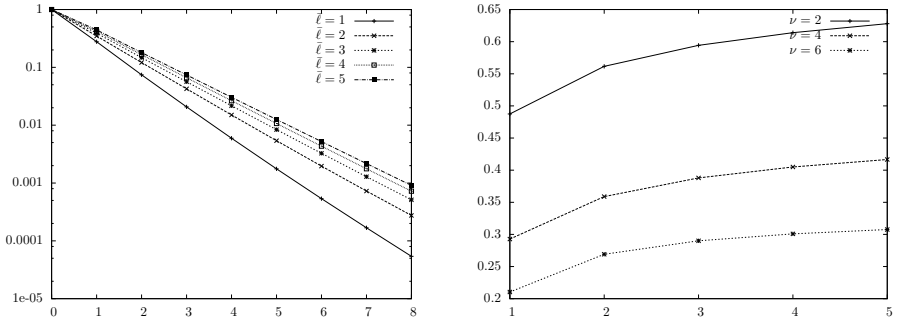


Figure 8. Convergence of multigrid V-cycle with damped Jacobi smoother. Left: r_k, for $k = 0, \ldots, 8$ and $\bar{\ell} = 1, \ldots, 5$. Right: $(r_8)^{\frac{1}{8}}$ for $\bar{\ell} = 1, \ldots, 5$ and $\nu = 2, 4, 6$.

For the method with damped Jacobi smoothing we also clearly observe an essentially level independent rate of convergence. Furthermore there is an increase in the rate of convergence when the number ν of smoothing step gets larger. Comparing the results of the multigrid method with Jacobi smoothing to those with SGS smoothing we see that the latter method has a significantly faster convergence. Note, however, that the arithmetic costs per iteration for the latter method are higher (the ratio lies between 1.5 and 2).

5 Multigrid Methods for Generalized Stokes Equations

Let $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3 be a bounded connected domain. We consider the following generalized Stokes problem: Given $\vec{f}$, find a velocity $\vec{u}$ and a pressure p such that

$$\begin{aligned} \xi\vec{u} - \nu\Delta\vec{u} + \nabla p &= \vec{f} \quad \text{in } \Omega, \\ \nabla \cdot \vec{u} &= 0 \quad \text{in } \Omega, \\ \vec{u} &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{37}$$

The parameters $\nu > 0$ (viscosity) and $\xi \geq 0$ are given. Often the latter is proportional to the inverse of the time step in an implicit time integration method applied to a nonstationary Stokes problem. Note that this general setting includes the classical (stationary) Stokes problem ($\xi = 0$). The weak formulation of (37) is as follows: Given $\vec{f} \in L^2(\Omega)^d$, we seek $\vec{u} \in H_0^1(\Omega)^d$ and $p \in L_0^2(\Omega) := \{\, q \in L^2(\Omega) \mid \int_\Omega q\, dx = 0 \,\}$ such that

$$\begin{aligned} \xi(\vec{u},\vec{v}) + \nu(\nabla\vec{u},\nabla\vec{v}) - (\text{div}\,\vec{v},p) &= (\vec{f},\vec{v}) \quad \text{for all } \vec{v} \in H_0^1(\Omega)^d, \\ (\text{div}\,\vec{u},q) &= 0 \quad \text{for all } q \in L_0^2(\Omega). \end{aligned} \tag{38}$$

Here $(\cdot,\cdot)$ denotes the L^2 scalar product.

For discretization of (38) we use a standard finite element approach. Based on a regular family of *nested* tetrahedral grids $\mathcal{T}_\ell = \mathcal{T}_{h_\ell}$ with $\mathcal{T}_0 \subset \mathcal{T}_1 \subset \ldots$ we use a sequence of nested finite element spaces

$$(\mathbf{V}_{\ell-1}, Q_{\ell-1}) \subset (\mathbf{V}_\ell, Q_\ell), \quad \ell = 1, 2, \ldots.$$

The pair of spaces $(\mathbf{V}_\ell, Q_\ell)$, $\ell \geq 0$, is assumed to be stable. By h_ℓ we denote the mesh size parameter corresponding to $\mathcal{T}_\ell$. In our numerical experiments we use the Hood-Taylor $\mathcal{P}_2 - \mathcal{P}_1$ pair:

$$\begin{aligned} &\mathbf{V}_\ell = V_\ell^d, \quad V_\ell := \{\, v \in C(\Omega) \mid v_{|T} \in \mathcal{P}_2 \quad \text{for all } T \in \mathcal{T}_\ell \,\}, \\ &Q_\ell = \{\, v \in C(\Omega) \mid v_{|T} \in \mathcal{P}_1 \quad \text{for all } T \in \mathcal{T}_\ell \,\}. \end{aligned} \tag{39}$$

The discrete problem is given by the Galerkin discretization of (38) with the pair $(\mathbf{V}_\ell, Q_\ell)$. We are interested in the solution of this discrete problem on a given finest discretization level $\ell = \bar{\ell}$. The resulting discrete problem can be represented using the standard nodal bases in these finite element spaces. The representation of the discrete problem on level ℓ in these bases results in a *linear saddle point problem* of the form

$$\mathcal{A}_\ell \mathbf{x}_\ell = \mathbf{b}_\ell \quad \text{with} \quad \mathcal{A}_\ell = \begin{pmatrix} A_\ell & B_\ell^T \\ B_\ell & 0 \end{pmatrix}, \quad \mathbf{x}_\ell = \begin{pmatrix} \mathbf{u}_\ell \\ \mathbf{p}_\ell \end{pmatrix}. \tag{40}$$

The dimensions of the spaces $\mathbf{V}_\ell$ and Q_ℓ are denoted by n_ℓ and m_ℓ, respectively. The matrix $A_\ell \in \mathbb{R}^{n_\ell \times n_\ell}$ is the discrete representation of the differential operator $\xi I - \nu\Delta$ and is symmetric positive definite. Note that A_ℓ depends on the parameters ξ and ν. The matrix $\mathcal{A}_\ell$ depends on these parameters, too, and is *symmetric and strongly indefinite*.

We describe a multigrid method that can be used for the iterative solution of the system (40). This method has the same algorithmic structure as in (31). We need intergrid transfer operators (prolongation and restriction) and a smoother. These components are described below.

Intergrid transfer operators. For the prolongation and restriction of vectors (or corresponding finite element functions) between different level we use the canonical operators. The prolongation between level $\ell - 1$ and ℓ is given by

$$P_\ell = \begin{pmatrix} P_{\mathbf{V}} & 0 \\ 0 & P_Q \end{pmatrix}, \tag{41}$$

where the matrices $P_{\mathbf{V}} : \mathbb{R}^{n_{\ell-1}} \to \mathbb{R}^{n_\ell}$ and $P_Q : \mathbb{R}^{m_{\ell-1}} \to \mathbb{R}^{m_\ell}$ are matrix representations of the embeddings $\mathbf{V}_{\ell-1} \subset \mathbf{V}_\ell$ (quadratic interpolation for $\mathcal{P}_2$) and $Q_{\ell-1} \subset Q_\ell$

(linear interpolation for $\mathcal{P}_1$), respectively. For the restriction operator R_ℓ between the levels ℓ and $\ell-1$ we take the adjoint of P_ℓ (w.r.t. a scaled Euclidean scalar product). Then the Galerkin property $\mathcal{A}_{\ell-1} = R_\ell \mathcal{A}_\ell P_\ell$ holds.

Braess-Sarazin smoother. This smoother is introduced in Braess[10]. With $D_\ell = \mathrm{diag}(A_\ell)$ and a given $\alpha > 0$ the smoothing iteration has the form

$$\begin{pmatrix} \mathbf{u}_\ell^{k+1} \\ \mathbf{p}_\ell^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{u}_\ell^{k} \\ \mathbf{p}_\ell^{k} \end{pmatrix} - \begin{pmatrix} \alpha D_\ell & B_\ell^T \\ B_\ell & 0 \end{pmatrix}^{-1} \left\{ \begin{pmatrix} A_\ell & B_\ell^T \\ B_\ell & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_\ell^{k} \\ \mathbf{p}_\ell^{k} \end{pmatrix} - \begin{pmatrix} \mathbf{f}_\ell \\ 0 \end{pmatrix} \right\}. \tag{42}$$

Each iteration (42) requires the solution of the auxiliary problem

$$\begin{pmatrix} \alpha D_\ell & B_\ell^T \\ B_\ell & 0 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{u}}_\ell \\ \hat{\mathbf{p}}_\ell \end{pmatrix} = \begin{pmatrix} \mathbf{r}_\ell^k \\ B_\ell \mathbf{u}_\ell^k \end{pmatrix} \tag{43}$$

with $\mathbf{r}_\ell^k = A_\ell \mathbf{u}_\ell^k + B_\ell^T \mathbf{p}_\ell^k - \mathbf{f}_\ell$. From (43) one obtains

$$B_\ell \hat{\mathbf{u}}_\ell = B_\ell \mathbf{u}_\ell^k,$$

and hence,

$$B_\ell \mathbf{u}_\ell^{k+1} = B_\ell(\mathbf{u}_\ell^k - \hat{\mathbf{u}}_\ell) = 0 \quad \text{for all} \quad j \geq 0. \tag{44}$$

Therefore, the Braess-Sarazin method can be considered as a smoother on the subspace of vectors that satisfy the constraint equation $B_\ell \mathbf{u}_\ell = 0$.

The problem (43) can be reduced to a problem for the auxiliary pressure unknown $\hat{\mathbf{p}}_\ell$:

$$Z_\ell \hat{\mathbf{p}}_\ell = B_\ell D_\ell^{-1} \mathbf{r}_\ell^k - \alpha B_\ell \mathbf{u}_\ell^k, \tag{45}$$

where $Z_\ell = B_\ell D_\ell^{-1} B_\ell^T$.

Remark 5.1 The matrix Z_ℓ is similar to a discrete Laplace operator on the pressure space. In practice the system (45) is solved approximately using an efficient iterative solver, cf. Braess[10], Zulehner[11]. □

Once $\hat{\mathbf{p}}_\ell$ is known (approximately), an approximation for $\hat{\mathbf{u}}_\ell$ can easily be determined from $\alpha D_\ell \hat{\mathbf{u}}_\ell = \mathbf{r}_\ell^k - B_\ell^T \hat{\mathbf{p}}_\ell$.

Vanka smoother. The Vanka-type smoothers, originally proposed by Vanka[12] for finite difference schemes, are block Gauß-Seidel type of methods. If one uses such a method in a finite element setting then a block of unknowns consists of all degrees of freedom that correspond with one element. Numerical tests given in John[13] show that the use of this element-wise Vanka smoother can be problematic for continuous pressure approximations. In John[13] the pressure-oriented Vanka smoother for continuous pressure approximations has been suggested as a good alternative. In this method a local problem corresponds to the block of unknowns consisting of one pressure unknown and all velocity degrees of freedom that are connected with this pressure unknown. We only consider this type of Vanka smoother. We first give a more precise description of this method.

We take a fixed level ℓ in the discretization hierarchy. To simplify the presentation we drop the level index ℓ from the notation, i.e. we write, for example, $\begin{pmatrix} \mathbf{u} \\ \mathbf{p} \end{pmatrix} \in \mathbb{R}^{n+m}$ instead of $\begin{pmatrix} \mathbf{u}_\ell \\ \mathbf{p}_\ell \end{pmatrix} \in \mathbb{R}^{n_\ell + m_\ell}$. Let $r_P^{(j)} : \mathbb{R}^m \to \mathbb{R}$ be the pressure projection (injection)

$$r_P^{(j)} \mathbf{p} = p_j, \qquad j = 1, \ldots, m.$$

For each j ($1 \leq j \leq m$) let the set of velocity indices that are "connected" to j be given by

$$\mathcal{V}_j = \{1 \leq i \leq n \mid (r_P^{(j)} B)_i \neq 0\}.$$

Define $d_j := |\mathcal{V}_j|$ and write $\mathcal{V}_j = \{i_1 < i_2 < \ldots < i_{d_j}\}$. A corresponding velocity projection operator $r_V^{(j)} : \mathbb{R}^n \to \mathbb{R}^{d_j}$ is given by

$$r_V^{(j)} \mathbf{u} = (u_{i_1}, u_{i_2}, \ldots, u_{i_{d_j}})^T.$$

The combined pressure and velocity projection is given by

$$r^{(j)} = \begin{pmatrix} r_V^{(j)} & 0 \\ 0 & r_P^{(j)} \end{pmatrix} \in \mathbb{R}^{(d_j+1)\times(n+m)}.$$

Furthermore, define $p^{(j)} = \left(r^{(j)}\right)^T$. Using these operators we can formulate a standard multiplicative Schwarz method. Define

$$\mathcal{A}^{(j)} := r^{(j)} \mathcal{A} p^{(j)} =: \begin{pmatrix} A^{(j)} & {B^{(j)}}^T \\ B^{(j)} & 0 \end{pmatrix} \in \mathbb{R}^{(d_j+1)\times(d_j+1)}.$$

Note that $B^{(j)}$ is a row vector of length d_j. In addition, we define

$$\mathcal{D}^{(j)} = \begin{pmatrix} \mathrm{diag}(A^{(j)}) & {B^{(j)}}^T \\ B^{(j)} & 0 \end{pmatrix} = \begin{pmatrix} \ddots & 0 & \vdots \\ 0 & \ddots & \vdots \\ \ldots & \ldots & 0 \end{pmatrix} \in \mathbb{R}^{(d_j+1)\times(d_j+1)}.$$

The *full* Vanka smoother is a multiplicative Schwarz method (or block Gauss-Seidel method) with iteration matrix

$$\mathcal{S}_{\text{full}} = \prod_{j=1}^{m} \left(I - p^{(j)} (\mathcal{A}^{(j)})^{-1} r^{(j)} \mathcal{A}\right). \tag{46}$$

The *diagonal* Vanka smoother is similar, but with $\mathcal{D}^{(j)}$ instead of $\mathcal{A}^{(j)}$:

$$\mathcal{S}_{\text{diag}} = \prod_{j=1}^{m} \left(I - p^{(j)} (\mathcal{D}^{(j)})^{-1} r^{(j)} \mathcal{A}\right). \tag{47}$$

Thus, a smoothing step with a Vanka-type smoother consists of a loop over all pressure degrees of freedom ($j = 1, \ldots, m$), where for each j a linear system of equations with the matrix $\mathcal{A}^{(j)}$ (or $\mathcal{D}^{(j)}$) has to be solved. The degrees of freedom are updated in a Gauss-Seidel manner. These two methods are well-defined if all matrices $\mathcal{A}^{(j)}$ and $\mathcal{D}^{(j)}$ are nonsingular.

The linear systems with the diagonal Vanka smoother can be solved very efficiently using the special structure of the matrix $\mathcal{D}^{(j)}$ whereas for the systems with the full Vanka smoother a direct solver for the systems with the matrices $\mathcal{A}^{(j)}$ is required. The computational costs for solving a local (i.e. for each block) linear system of equations is $\sim d_j$ for the diagonal Vanka smoother and $\sim d_j^3$ for the full Vanka smoother. Typical values for d_j are given in Table 2.

Using the prolongation, restriction and smoothers as explained above a multigrid algorithm for solving the discretized Stokes problem (40) is defined as in (31).

	$h_0 = 2^{-1}$	$h_1 = 2^{-2}$	$h_2 = 2^{-3}$	$h_3 = 2^{-4}$	$h_4 = 2^{-5}$
n_ℓ	81	1029	10125	89373	750141
m_ℓ	27	125	729	4913	35937

Table 1. Dimensions: n_ℓ = number of velocity unknowns, m_ℓ = number of pressure unknowns.

6 Numerical Experiment: Multigrid Applied to a Generalized Stokes Equation

We consider the generalized Stokes equation as in (37) on the unit cube $\Omega = (0,1)^3$. The right-hand side $\vec{f}$ is taken such that the continuous solution is

$$\vec{u}(x,y,z) = \frac{1}{3}\begin{pmatrix} \sin(\pi x)\sin(\pi y)\sin(\pi z) \\ -\cos(\pi x)\cos(\pi y)\sin(\pi z) \\ 2\cdot\cos(\pi x)\sin(\pi y)\cos(\pi z)\end{pmatrix},$$

$$p(x,y,z) = \cos(\pi x)\sin(\pi y)\sin(\pi z) + C$$

with a constant C such that $\int_\Omega p\,dx = 0$. For the discretization we start with a uniform tetrahedral grid with $h_0 = \frac{1}{2}$ and we apply regular refinements to this starting discretization. For the finite element discretization we use the Hood-Taylor $\mathcal{P}_2$-$\mathcal{P}_1$ pair, cf. (39). In Table 1 the dimension of the system to be solved on each level and the corresponding mesh size are given.

In all tests below the iterations were repeated until the condition

$$\frac{\|\mathbf{r}^{(k)}\|}{\|\mathbf{r}^{(0)}\|} < 10^{-10},$$

with $\mathbf{r}^{(k)} = \mathbf{b} - \mathcal{A}\mathbf{x}^{(k)}$, was satisfied.

We first consider an experiment to show that for this problem class the multigrid method with *full* Vanka smoother is very time consuming. In Table 2 we show the maximal and mean values of d_j on the level ℓ. These numbers indicate the dimensions of the local systems that have to be solved in the Vanka smoother.

	$h_0 = 2^{-1}$	$h_1 = 2^{-2}$	$h_2 = 2^{-3}$	$h_3 = 2^{-4}$	$h_4 = 2^{-5}$
$\frac{\text{mean}(d_j)}{\max_j d_j}$	21.8 / 82	51.7 / 157	88.8 / 157	119.1 / 165	138.1 / 166

Table 2. The maximal and mean values of d_j on different grids.

We use a multigrid W-cycle with 2 pre- and 2 post-smoothing iterations. In Table 3 we show the computing time (in seconds) and the number of iterations needed both for the full Vanka $\mathcal{S}_{\text{full}}$ and the diagonal Vanka $\mathcal{S}_{\text{diag}}$ smoother.

As can be seen from these results, the rather high dimensions of the local systems lead to high computing times for the multigrid method with the full Vanka smoother compared to the method with the diagonal Vanka smoother. Therefore we prefer the method with

$\xi = 0$	$\mathcal{S}_{\text{full}}$, $h_3 = 2^{-4}$	$\mathcal{S}_{\text{diag}}$,$h_3 = 2^{-4}$	$\mathcal{S}_{\text{full}}$, $h_4 = 2^{-5}$	$\mathcal{S}_{\text{diag}}$,$h_4 = 2^{-5}$
$\nu = 1$	287 (4)	19 (10)	3504 (5)	224 (13)
$\nu = 10^{-1}$	283 (4)	19 (10)	3449 (5)	238 (13)
$\nu = 10^{-2}$	284 (4)	19 (10)	3463 (5)	238 (13)
$\nu = 10^{-3}$	356 (5)	20 (11)	3502 (5)	238 (13)

Table 3. CPU time and number of iterations for multigrid with the full and the diagonal Vanka smoother.

the diagonal Vanka smoother. In numerical experiments we observed that the multigrid W-cycle with only *one* pre- and post-smoothing iteration with the diagonal Vanka method sometimes diverges. Further tests indicate that often for the method with diagonal Vanka smoothing the choice $\nu_1 = \nu_2 = 4$ is (slightly) better (w.r.t. CPU time) than $\nu_1 = \nu_2 = 2$.

Results for two variants of the multigrid W-cycle method, one with diagonal Vanka smoothing (V-MGM) and one with Braess-Sarazin smoothing (BS-MGM) are given in the tables 4 and 5. In the V-MGM we use $\nu_1 = \nu_2 = 4$. Based on numerical experiments, in the method with the Braess-Sarazin smoother we use $\nu_1 = \nu_2 = 2$ and $\alpha = 1.25$. For other values $\alpha \in [1.1, 1.5]$ the efficiency is very similar. The linear system in (45) is solved approximately using a conjugate gradient method with a fixed relative tolerance $\varepsilon_{CG} = 10^{-2}$. To investigate the robustness of these method we give results for several values of ℓ, ν and ξ.

$\xi = 0$	$h_3 = 2^{-4}$	
ν	V-MGM	BS-MGM
$\nu = 1$	19 (5)	20 (11)
$\nu = 10^{-1}$	19 (5)	20 (11)
$\nu = 10^{-3}$	19 (5)	17 (8)
$\xi = 10$	$h_3 = 2^{-4}$	
ν	V-MGM	BS-MGM
$\nu = 1$	19 (5)	20 (11)
$\nu = 10^{-1}$	17 (4)	20 (11)
$\nu = 10^{-3}$	15 (3)	21 (7)
$\xi = 100$	$h_3 = 2^{-4}$	
ν	V-MGM	BS-MGM
$\nu = 1$	17 (4)	20 (11)
$\nu = 10^{-1}$	15 (3)	19 (7)
$\nu = 10^{-3}$	15 (3)	19 (6)

Table 4. CPU time and the number of iterations for BS- and V-MGM methods.

The results show that the rate of convergence is essentially independent of the parameters ν and ξ, i.e., these methods have a robustness property. Furthermore we observe that if for fixed ν, ξ we compare the results for $\ell = 3$ ($h_3 = 2^{-4}$) with those for $\ell = 4$ ($h_4 = 2^{-5}$) then for the V-MGM there is (almost) no increase in the number of iterations. This illustrates the mesh independent rate of convergence of the method. For the BS-MGM there is a (small) growth in the number of iterations. For both methods the CPU

$\xi = 0$	$h_4 = 2^{-5}$	
ν	V-MGM	BS-MGM
$\nu = 1$	198 (5)	274 (14)
$\nu = 10^{-1}$	199 (5)	276 (14)
$\nu = 10^{-3}$	198 (5)	241 (11)
$\xi = 10$	$h_3 = 2^{-5}$	
ν	V-MGM	BS-MGM
$\nu = 1$	190 (5)	244 (13)
$\nu = 10^{-1}$	189 (5)	224 (10)
$\nu = 10^{-3}$	145 (3)	238 (7)
$\xi = 100$	$h_3 = 2^{-5}$	
ν	V-MGM	BS-MGM
$\nu = 1$	190 (5)	241 (13)
$\nu = 10^{-1}$	167 (4)	243 (13)
$\nu = 10^{-3}$	122 (2)	282 (9)

Table 5. CPU time and the number of iterations for BS- and V-MGM methods.

time needed per iteration grows with a factor of roughly 10 when going from $\ell = 3$ to $\ell = 4$. The number of unknowns then grows with about a factor 8.3, cf. Table 1. This indicates that the arithmetic work per iteration is almost linear in the number of unknowns.

7 Convergence Analysis for Scalar Elliptic Problems

In this section we present a convergence analysis for the multigrid method introduced in section 3. Our approach is based on the so-called approximation- and smoothing property, introduced by Hackbusch[1, 14]. For a discussion of other analyses we refer to remark 7.23.

7.1 Introduction

One easily verifies that the two-grid method is a linear iterative method. The iteration matrix of this method with ν_1 presmoothing and ν_2 postsmoothing iterations on level ℓ is given by

$$\mathbf{C}_{TG,\ell} = \mathbf{C}_{TG,\ell}(\nu_2, \nu_1) = \mathbf{S}_\ell^{\nu_2}(\mathbf{I} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1} \mathbf{r}_\ell \mathbf{A}_\ell)\mathbf{S}_\ell^{\nu_1} \tag{48}$$

with $\mathbf{S}_\ell = \mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell$ the iteration matrix of the smoother.

Theorem 7.1 *The multigrid method* (31) *is a linear iterative method with iteration matrix* $\mathbf{C}_{MG,\ell}$ *given by*

$$\mathbf{C}_{MG,0} = 0 \tag{49a}$$

$$\mathbf{C}_{MG,\ell} = \mathbf{S}_\ell^{\nu_2}\big(\mathbf{I} - \mathbf{p}_\ell(\mathbf{I} - \mathbf{C}_{MG,\ell-1}^{\tau})\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\big)\mathbf{S}_\ell^{\nu_1} \tag{49b}$$

$$= \mathbf{C}_{TG,\ell} + \mathbf{S}_\ell^{\nu_2}\mathbf{p}_\ell\mathbf{C}_{MG,\ell-1}^{\tau}\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{S}_\ell^{\nu_1}, \quad \ell = 1, 2, \ldots \tag{49c}$$

Proof: The result in (49a) is trivial. The result in (49c) follows from (49b) and the definition of $\mathbf{C}_{TG,\ell}$. We now prove the result in (49b) by induction. For $\ell = 1$ it follows from (49a) and (48). Assume that the result is correct for $\ell - 1$. Then $\mathrm{MGM}_{\ell-1}(\mathbf{y}_{\ell-1}, \mathbf{z}_{\ell-1})$ defines a linear iterative method and for arbitrary $\mathbf{y}_{\ell-1},\ \mathbf{z}_{\ell-1} \in \mathbb{R}^{n_{\ell-1}}$ we have

$$\mathrm{MGM}_{\ell-1}(\mathbf{y}_{\ell-1}, \mathbf{z}_{\ell-1}) - \mathbf{A}_{\ell-1}^{-1}\mathbf{z}_{\ell-1} = \mathbf{C}_{MG,\ell-1}(\mathbf{y}_{\ell-1} - \mathbf{A}_{\ell-1}^{-1}\mathbf{z}_{\ell-1}) \tag{50}$$

We rewrite the algorithm (31) as follows:

$$\begin{aligned}
\mathbf{x}^1 &:= \mathcal{S}_\ell^{\nu_1}(\mathbf{x}_\ell^{\mathrm{old}}, \mathbf{b}_\ell) \\
\mathbf{x}^2 &:= \mathbf{x}^1 + \mathbf{p}_\ell \mathrm{MGM}_{\ell-1}^{\tau}\big(0, \mathbf{r}_\ell(\mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}^1)\big) \\
\mathbf{x}_\ell^{\mathrm{new}} &:= \mathcal{S}_\ell^{\nu_2}(\mathbf{x}^2, \mathbf{b}_\ell).
\end{aligned}$$

From this we get

$$\begin{aligned}
\mathbf{x}_\ell^{\mathrm{new}} - \mathbf{x}_\ell^* &= \mathbf{x}_\ell^{\mathrm{new}} - \mathbf{A}_\ell^{-1}\mathbf{b}_\ell = \mathbf{S}_\ell^{\nu_2}(\mathbf{x}^2 - \mathbf{x}_\ell^*) \\
&= \mathbf{S}_\ell^{\nu_2}\big(\mathbf{x}^1 - \mathbf{x}_\ell^* + \mathbf{p}_\ell \mathrm{MGM}_{\ell-1}^{\tau}\big(0, \mathbf{r}_\ell(\mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}^1)\big)\big).
\end{aligned}$$

Now we use the result (50) with $\mathbf{y}_{\ell-1} = 0,\ \mathbf{z}_{\ell-1} := \mathbf{r}_\ell(\mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}^1)$. This yields

$$\begin{aligned}
\mathbf{x}_\ell^{\mathrm{new}} - \mathbf{x}_\ell^* &= \mathbf{S}_\ell^{\nu_2}\big(\mathbf{x}^1 - \mathbf{x}_\ell^* + \mathbf{p}_\ell(\mathbf{A}_{\ell-1}^{-1}\mathbf{z}_{\ell-1} - \mathbf{C}_{MG,\ell-1}^{\tau}\mathbf{A}_{\ell-1}^{-1}\mathbf{z}_{\ell-1}\big) \\
&= \mathbf{S}_\ell^{\nu_2}\big(\mathbf{I} - \mathbf{p}_\ell(\mathbf{I} - \mathbf{C}_{MG,\ell-1}^{\tau})\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\big)(\mathbf{x}^1 - \mathbf{x}_\ell^*) \\
&= \mathbf{S}_\ell^{\nu_2}\big(\mathbf{I} - \mathbf{p}_\ell(\mathbf{I} - \mathbf{C}_{MG,\ell-1}^{\tau})\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\big)\mathbf{S}_\ell^{\nu_1}(\mathbf{x}^{\mathrm{old}} - \mathbf{x}_\ell^*).
\end{aligned}$$

This completes the proof. ∎

The convergence analysis will be based on the following splitting of the two-grid iteration matrix, with $\nu_2 = 0$, i.e. no postsmoothing:

$$\begin{aligned}
\|\mathbf{C}_{TG,\ell}(0,\nu_1)\|_2 &= \|(\mathbf{I} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell)\mathbf{S}_\ell^{\nu_1}\|_2 \\
&\le \|\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\|_2\,\|\mathbf{A}_\ell\mathbf{S}_\ell^{\nu_1}\|_2
\end{aligned} \tag{51}$$

In section 7.2 we will prove a bound of the form $\|\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\|_2 \le C_A\|\mathbf{A}_\ell\|_2^{-1}$. This result is called the *approximation property*. In section 7.3 we derive a suitable bound for the term $\|\mathbf{A}_\ell\mathbf{S}_\ell^{\nu_1}\|_2$. This is the so-called *smoothing property*. In section 7.4 we combine these bounds with the results in (51) and in theorem 7.1. This yields bounds for the contraction number of the two-grid method and of the multigrid W-cycle. For the V-cycle a more subtle analysis is needed. This is presented in section 7.5. In the convergence analysis we need the following:

Assumption 7.2 *In the sections 7.2–7.5 we assume that the family of triangulations $\{\mathcal{T}_{h_\ell}\}$ corresponding to the finite element spaces V_ℓ, $\ell = 0, 1, \ldots$, is* quasi-uniform *and that $h_{\ell-1} \le c h_\ell$ with a constant c independent of ℓ.*

We give some results that will be used in the analysis further on. First we recall an *inverse inequality* that is known from the analysis of finite element methods:

$$|v_\ell|_1 \le c\, h_\ell^{-1}\|v_\ell\|_{L^2} \quad \text{for all } v_\ell \in V_\ell$$

with a constant c independent of ℓ. For this result to hold we need assumption 7.2.
We now show that, apart from a scaling factor, the isomorphism $P_\ell : (\mathbb{R}^{n_\ell}, \langle\cdot,\cdot\rangle) \to (V_\ell, \langle\cdot,\cdot\rangle_{L^2})$ and its inverse are uniformly (w.r.t. ℓ) bounded:

Lemma 7.3 *There exist constants $c_1 > 0$ and c_2 independent of ℓ such that*

$$c_1 \|P_\ell \mathbf{x}\|_{L^2} \le h_\ell^{\frac{1}{2}d} \|\mathbf{x}\|_2 \le c_2 \|P_\ell \mathbf{x}\|_{L^2} \quad \textit{for all } \mathbf{x} \in \mathbb{R}^{n_\ell}. \tag{52}$$

Proof: The definition of P_ℓ yields $P_\ell \mathbf{x} = \sum_{i=1}^{n_\ell} x_i \phi_i =: v_\ell \in V_\ell$ and $v_\ell(\xi_i) = x_i$, where ξ_i is the vertex in the triangulation which corresponds to the nodal basis function ϕ_i. Note that

$$\|P_\ell \mathbf{x}\|_{L^2}^2 = \|v_\ell\|_{L^2}^2 = \sum_{T \in \mathcal{T}_\ell} \|v_\ell\|_{L^2(T)}^2.$$

Since v_ℓ is linear on each simplex T in the triangulation $\mathcal{T}_\ell$ there are constants $\tilde{c}_1 > 0$ and $\tilde{c}_2$ independent of h_ℓ such that

$$\tilde{c}_1 \|v_\ell\|_{L^2(T)}^2 \le |T| \sum_{\xi_j \in V(T)} v_\ell(\xi_j)^2 \le \tilde{c}_2 \|v_\ell\|_{L^2(T)}^2,$$

where $V(T)$ denotes the set of vertices of the simplex T. Summation over all $T \in \mathcal{T}_\ell$, using $v_\ell(\xi_j) = x_j$ and $|T| \sim h_\ell^d$ we obtain

$$\hat{c}_1 \|v_\ell\|_{L^2}^2 \le h_\ell^d \sum_{i=1}^{n_\ell} x_i^2 \le \hat{c}_2 \|v_\ell\|_{L^2}^2,$$

with constants $\hat{c}_1 > 0$ and $\hat{c}_2$ independent of h_ℓ and thus we get the result in (52). ∎

The third preliminary result concerns the scaling of the stiffness matrix:

Lemma 7.4 *Let $\mathbf{A}_\ell$ be the stiffness matrix as in* (26)*. Assume that the bilinear form is such that the usual conditions* (22) *are satisfied. Then there exist constants $c_1 > 0$ and c_2 independent of ℓ such that*

$$c_1 h_\ell^{d-2} \le \|\mathbf{A}_\ell\|_2 \le c_2 h_\ell^{d-2}.$$

Proof: First note that

$$\|\mathbf{A}_\ell\|_2 = \max_{\mathbf{x},\mathbf{y} \in \mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}.$$

Using the result in lemma 7.3, the continuity of the bilinear form and the inverse inequality we get

$$\begin{aligned} \max_{\mathbf{x},\mathbf{y} \in \mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} &\le c h_\ell^d \max_{v_\ell, w_\ell \in V_\ell} \frac{k(v_\ell, w_\ell)}{\|v_\ell\|_{L^2} \|w_\ell\|_{L^2}} \\ &\le c h_\ell^d \max_{v_\ell, w_\ell \in V_\ell} \frac{|v_\ell|_1 |w_\ell|_1}{\|v_\ell\|_{L^2} \|w_\ell\|_{L^2}} \le c\, h_\ell^{d-2} \end{aligned}$$

and thus the upper bound is proved. The lower bound follows from

$$\max_{\mathbf{x},\mathbf{y} \in \mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \ge \max_{1 \le i \le n_\ell} \langle \mathbf{A}_\ell \mathbf{e}_i, \mathbf{e}_i \rangle = k(\phi_i, \phi_i) \ge c |\phi_i|_1^2 \ge c h_\ell^{d-2}$$

The last inequality can be shown by using for $T \subset \operatorname{supp}(\phi_i)$ the affine transformation from the unit simplex to T. ∎

7.2 Approximation property

In this section we derive a bound for the first factor in the splitting (51). We start with two important assumptions that are crucial for the analysis. This first one concerns *regularity of the continuous problem*, the second one is a *discretization error bound*.

Assumption 7.5 *We assume that the continuous problem in* (23) *is H^2-regular, i.e. for $f \in L^2(\Omega)$ the corresponding solution u satisfies*

$$\|u\|_{H^2} \leq c\,\|f\|_{L^2},$$

with a constant c independent of f. Furthermore we assume a finite element discretization error bound for the Galerkin discretization (25)*:*

$$\|u - u_\ell\|_{L^2} \leq c h_\ell^2 \|f\|_{L^2}$$

with c independent of f and of ℓ.

We will need the *dual* problem of (23) which is as follows: determine $\tilde{u} \in H_0^1(\Omega)$ such that $k(v, \tilde{u}) = f(v)$ for all $v \in H_0^1(\Omega)$. Note that this dual problem is obtained by interchanging the arguments in the bilinear form $k(\cdot,\cdot)$ and that the dual problem equals the original one if the bilinear form is symmetric (as for example in case of the Poisson equation).

In the analysis we will use the adjoint operator $P_\ell^* : V_\ell \to \mathbb{R}^{n_\ell}$ which satisfies $\langle P_\ell \mathbf{x}, v_\ell\rangle_{L^2} = \langle \mathbf{x}, P_\ell^* v_\ell\rangle$ for all $\mathbf{x} \in \mathbb{R}^{n_\ell},\ v_\ell \in V_\ell$. As a direct consequence of lemma 7.3 we obtain

$$c_1 \|P_\ell^* v_\ell\|_2 \leq h_\ell^{\frac{1}{2}d} \|v_\ell\|_{L^2} \leq c_2 \|P_\ell^* v_\ell\|_2 \quad \text{for all } \ v_\ell \in V_\ell \tag{53}$$

with constants $c_1 > 0$ and c_2 independent of ℓ. We now formulate a main result for the convergence analysis of multigrid methods:

Theorem 7.6 (Approximation property.) *Consider $\mathbf{A}_\ell,\ \mathbf{p}_\ell,\ \mathbf{r}_\ell$ as defined in* (26), (29),(30)*. Assume that the variational problem* (23) *is such that the usual conditions* (22) *are satisfied. Moreover, the problem* (23) *and the corresponding dual problem are assumed to be H^2-regular. Then there exists a constant C_A independent of ℓ such that*

$$\|\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1} \mathbf{r}_\ell\|_2 \leq C_A \|\mathbf{A}_\ell\|_2^{-1} \quad \textit{for } \ \ell = 1, 2, \ldots \tag{54}$$

Proof: Let $\mathbf{b}_\ell \in \mathbb{R}^{n_\ell}$ be given. The constants in the proof are independent of $\mathbf{b}_\ell$ and of ℓ. Consider the variational problems:

$$\begin{aligned}
u \in H_0^1(\Omega): &\quad k(u, v) = \langle (P_\ell^*)^{-1} \mathbf{b}_\ell, v\rangle_{L^2} \quad \text{for all } \ v \in H_0^1(\Omega) \\
u_\ell \in V_\ell: &\quad k(u_\ell, v_\ell) = \langle (P_\ell^*)^{-1} \mathbf{b}_\ell, v_\ell\rangle_{L^2} \quad \text{for all } \ v_\ell \in V_\ell \\
u_{\ell-1} \in V_{\ell-1}: &\quad k(u_{\ell-1}, v_{\ell-1}) = \langle (P_\ell^*)^{-1} \mathbf{b}_\ell, v_{\ell-1}\rangle_{L^2} \quad \text{for all } \ v_{\ell-1} \in V_{\ell-1}.
\end{aligned}$$

Then

$$\mathbf{A}_\ell^{-1} \mathbf{b}_\ell = P_\ell^{-1} u_\ell \ \text{ and } \ \mathbf{A}_{\ell-1}^{-1} \mathbf{r}_\ell \mathbf{b}_\ell = P_{\ell-1}^{-1} u_{\ell-1}$$

hold. Hence we obtain, using lemma 7.3,

$$\|(\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1} \mathbf{r}_\ell)\mathbf{b}_\ell\|_2 = \|P_\ell^{-1}(u_\ell - u_{\ell-1})\|_2 \leq c\, h_\ell^{-\frac{1}{2}d} \|u_\ell - u_{\ell-1}\|_{L^2}. \tag{55}$$

Now we use the assumptions on the discretization error bound and on the H^2-regularity of the problem. This yields

$$\begin{aligned} \|u_\ell - u_{\ell-1}\|_{L^2} &\leq \|u_\ell - u\|_{L^2} + \|u_{\ell-1} - u\|_{L^2} \\ &\leq ch_\ell^2 |u|_2 + + ch_{\ell-1}^2 |u|_2 \leq ch_\ell^2 \|(P_\ell^*)^{-1}\mathbf{b}_\ell\|_{L^2} \end{aligned} \tag{56}$$

We combine (55) with (56) and use (53), and get

$$\|(\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell)\mathbf{b}_\ell\|_2 \leq c\, h_\ell^{2-d} \|\mathbf{b}_\ell\|_2$$

and thus $\|\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell \mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\|_2 \leq c\, h_\ell^{2-d}$. The proof is completed if we use lemma 7.4. ■

Note that in the proof of the approximation property we use the underlying continuous problem.

7.3 Smoothing property

In this section we derive inequalities of the form

$$\|\mathbf{A}_\ell \mathbf{S}_\ell^\nu\|_2 \leq g(\nu)\|\mathbf{A}_\ell\|_2$$

where $g(\nu)$ is a monotonically decreasing function with $\lim_{\nu\to\infty} g(\nu) = 0$. In the first part of this section we derive results for the case that $\mathbf{A}_\ell$ is symmetric positive definite. In the second part we discuss the general case.

Smoothing property for the symmetric positive definite case
We start with an elementary lemma:
Lemma 7.7 *Let $\mathbf{B} \in \mathbb{R}^{m\times m}$ be a symmetric positive definite matrix with $\sigma(\mathbf{B}) \subset (0,1]$. Then we have*

$$\|\mathbf{B}(\mathbf{I} - \mathbf{B})^\nu\|_2 \leq \frac{1}{2(\nu+1)} \quad \textit{for } \nu = 1, 2, \ldots$$

Proof: Note that

$$\|\mathbf{B}(\mathbf{I} - \mathbf{B})^\nu\|_2 = \max_{x\in(0,1]} x(1-x)^\nu = \frac{1}{\nu+1}\Big(\frac{\nu}{\nu+1}\Big)^\nu.$$

A simple computation shows that $\nu \to \big(\frac{\nu}{\nu+1}\big)^\nu$ is decreasing on $[1,\infty)$. ■

Below for a few basic iterative methods we derive the smoothing property for the symmetric case, i.e., $\mathbf{b} = 0$ in the bilinear form $k(\cdot,\cdot)$. We first consider the Richardson method:
Theorem 7.8 *Assume that in the bilinear form we have $\mathbf{b} = 0$ and that the usual conditions* (22) *are satisfied. Let $\mathbf{A}_\ell$ be the stiffness matrix in* (26)*. For $c_0 \in (0,1]$ we have the smoothing property*

$$\|\mathbf{A}_\ell(\mathbf{I} - \frac{c_0}{\rho(\mathbf{A}_\ell)}\mathbf{A}_\ell)^\nu\|_2 \leq \frac{1}{2c_0(\nu+1)}\|\mathbf{A}_\ell\|_2\,, \quad \nu = 1, 2, \ldots$$

holds.

Proof: Note that $\mathbf{A}_\ell$ is symmetric positive definite. Apply lemma 7.7 with $\mathbf{B} := \omega_\ell \mathbf{A}_\ell$, $\omega_\ell := c_0\, \rho(\mathbf{A}_\ell)^{-1}$. This yields

$$\|\mathbf{A}_\ell(\mathbf{I} - \omega_\ell \mathbf{A}_\ell)^\nu\|_2 \leq \omega_\ell^{-1} \frac{1}{2(\nu+1)} \leq \frac{1}{2c_0(\nu+1)} \rho(\mathbf{A}_\ell) = \frac{1}{2c_0(\nu+1)} \|\mathbf{A}_\ell\|_2$$

and thus the result is proved. ∎

A similar result can be shown for the damped Jacobi method:

Theorem 7.9 *Assume that in the bilinear form we have* $\mathbf{b} = 0$ *and that the usual conditions* (22) *are satisfied. Let* $\mathbf{A}_\ell$ *be the stiffness matrix in* (26) *and* $\mathbf{D}_\ell := \mathrm{diag}(\mathbf{A}_\ell)$. *There exists an* $\omega \in (0, \rho(\mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^{-1}]$, *independent of* ℓ, *such that the smoothing property*

$$\|\mathbf{A}_\ell(\mathbf{I} - \omega \mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_2 \leq \frac{1}{2\omega(\nu+1)} \|\mathbf{A}_\ell\|_2\,, \quad \nu = 1, 2, \ldots$$

holds.

Proof: Define the symmetric positive definite matrix $\tilde{\mathbf{B}} := \mathbf{D}_\ell^{-\frac{1}{2}} \mathbf{A}_\ell \mathbf{D}_\ell^{-\frac{1}{2}}$. Note that

$$(D_\ell)_{ii} = (A_\ell)_{ii} = k(\phi_i, \phi_i) \geq c\, |\phi_i|_1^2 \geq c\, h_\ell^{d-2}\,, \tag{57}$$

with $c > 0$ independent of ℓ and i. Using this in combination with lemma 7.4 we get

$$\|\tilde{\mathbf{B}}\|_2 \leq \frac{\|\mathbf{A}_\ell\|_2}{\lambda_{\min}(\mathbf{D}_\ell)} \leq c\,, \quad c \text{ independent of } \ell.$$

Hence for $\omega \in (0, \frac{1}{c}] \subset (0, \rho(\mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^{-1}]$ we have $\sigma(\omega\tilde{\mathbf{B}}) \subset (0, 1]$. Application of lemma 7.7, with $\mathbf{B} = \omega\tilde{\mathbf{B}}$, yields

$$\begin{aligned}\|\mathbf{A}_\ell(\mathbf{I} - \omega \mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_2 &\leq \omega^{-1} \|\mathbf{D}_\ell^{\frac{1}{2}}\|_2 \|\omega\tilde{\mathbf{B}}(\mathbf{I} - \omega\tilde{\mathbf{B}})^\nu\|_2 \|\mathbf{D}_\ell^{\frac{1}{2}}\|_2 \\ &\leq \frac{\|\mathbf{D}_\ell\|_2}{2\omega(\nu+1)} \leq \frac{1}{2\omega(\nu+1)} \|\mathbf{A}_\ell\|_2\end{aligned}$$

and thus the result is proved. ∎

Remark 7.10 The value of the parameter ω used in theorem 7.9 is such that $\omega\rho(\mathbf{D}_\ell^{-1}\mathbf{A}_\ell) = \omega\rho(\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}) \leq 1$ holds. Note that

$$\rho(\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}) = \max_{\mathbf{x}\in\mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell\mathbf{x}, \mathbf{x}\rangle}{\langle \mathbf{D}_\ell\mathbf{x}, \mathbf{x}\rangle} \geq \max_{1\leq i\leq n_\ell} \frac{\langle \mathbf{A}_\ell\mathbf{e}_i, \mathbf{e}_i\rangle}{\langle \mathbf{D}_\ell\mathbf{e}_i\mathbf{e}_i\rangle} = 1$$

and thus we have $\omega \leq 1$. This explains why in multigrid methods one usually uses a *damped* Jacobi method as a smoother. □

We finally consider the symmetric Gauss-Seidel method. If $\mathbf{A}_\ell = \mathbf{A}_\ell^T$ this method has an iteration matrix

$$\mathbf{S}_\ell = \mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell, \quad \mathbf{M}_\ell = (\mathbf{D}_\ell - \mathbf{L}_\ell)\mathbf{D}_\ell^{-1}(\mathbf{D}_\ell - \mathbf{L}_\ell^T)\,, \tag{58}$$

where we use the decomposition $\mathbf{A}_\ell = \mathbf{D}_\ell - \mathbf{L}_\ell - \mathbf{L}_\ell^T$ with $\mathbf{D}_\ell$ a diagonal matrix and $\mathbf{L}_\ell$ a strictly lower triangular matrix.

Theorem 7.11 *Assume that in the bilinear form we have* $\mathbf{b} = 0$ *and that the usual conditions* (22) *are satisfied. Let* $\mathbf{A}_\ell$ *be the stiffness matrix in* (26) *and* $\mathbf{M}_\ell$ *as in* (58)*. The smoothing property*

$$\|\mathbf{A}_\ell(\mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_2 \leq \frac{c}{\nu+1}\|\mathbf{A}_\ell\|_2\,, \quad \nu = 1, 2, \ldots$$

holds with a constant c *independent of* ν *and* ℓ*.*

Proof: Note that $\mathbf{M}_\ell = \mathbf{A}_\ell + \mathbf{L}_\ell\mathbf{D}_\ell^{-1}\mathbf{L}_\ell^T$ and thus $\mathbf{M}_\ell$ is symmetric positive definite. Define the symmetric positive definite matrix $\mathbf{B} := \mathbf{M}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{M}_\ell^{-\frac{1}{2}}$. From

$$0 < \max_{\mathbf{x}\in\mathbb{R}^{n_\ell}} \frac{\langle \mathbf{B}\mathbf{x}, \mathbf{x}\rangle}{\langle \mathbf{x}, \mathbf{x}\rangle} = \max_{\mathbf{x}\in\mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell\mathbf{x}, \mathbf{x}\rangle}{\langle \mathbf{M}_\ell\mathbf{x}, \mathbf{x}\rangle} = \max_{\mathbf{x}\in\mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell\mathbf{x}, \mathbf{x}\rangle}{\langle \mathbf{A}_\ell\mathbf{x}, \mathbf{x}\rangle + \langle \mathbf{D}_\ell^{-1}\mathbf{L}_\ell^T\mathbf{x}, \mathbf{L}_\ell^T\mathbf{x}\rangle} \leq 1$$

it follows that $\sigma(\mathbf{B}) \subset (0, 1]$. Application of lemma 7.7 yields

$$\|\mathbf{A}_\ell(\mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_2 \leq \|\mathbf{M}_\ell^{\frac{1}{2}}\|_2^2\,\|\mathbf{B}(\mathbf{I} - \mathbf{B})^\nu\|_2 \leq \|\mathbf{M}_\ell\|_2 \frac{1}{2(\nu+1)}.$$

From (57) we have $\|\mathbf{D}_\ell^{-1}\|_2 \leq c\,h_\ell^{2-d}$. Using the sparsity of $\mathbf{A}_\ell$ we obtain

$$\|\mathbf{L}_\ell\|_2\|\mathbf{L}_\ell^T\|_2 \leq \|\mathbf{L}_\ell\|_\infty\|\mathbf{L}_\ell\|_1 \leq c(\max_{i,j}|(A_\ell)_{ij}|)^2 \leq c\|\mathbf{A}_\ell\|_2^2.$$

In combination with lemma 7.4 we then get

$$\|\mathbf{M}_\ell\|_2 \leq \|\mathbf{D}_\ell^{-1}\|_2\|\mathbf{L}_\ell\|_2\|\mathbf{L}_\ell^T\|_2 \leq c\,h_\ell^{2-d}\|\mathbf{A}_\ell\|_2^2 \leq c\|\mathbf{A}_\ell\|_2 \tag{59}$$

and this completes the proof. ∎

For the symmetric positive definite case smoothing properties have also been proved for other iterative methods. For example, in Wittum[15,16] a smoothing property is proved for a variant of the ILU method and in Bröker et al.[17] it is shown that the SPAI (sparse approximate inverse) preconditioner satisfies a smoothing property.

Smoothing property for the nonsymmetric case

For the analysis of the smoothing property in the general (possibly nonsymmetric) case we can not use lemma 7.7. Instead the analysis will be based on the following lemma (cf. Reusken[18,19]):

Lemma 7.12 *Let* $\|\cdot\|$ *be any induced matrix norm and assume that for* $\mathbf{B} \in \mathbb{R}^{m\times m}$ *the inequality* $\|\mathbf{B}\| \leq 1$ *holds. The we have*

$$\|(\mathbf{I} - \mathbf{B})(\mathbf{I} + \mathbf{B})^\nu\| \leq 2^{\nu+1}\sqrt{\frac{2}{\pi\nu}}\,, \quad \textit{for } \nu = 1, 2, \ldots$$

Proof: Note that

$$(\mathbf{I} - \mathbf{B})(\mathbf{I} + \mathbf{B})^\nu = (\mathbf{I} - \mathbf{B})\sum_{k=0}^{\nu}\binom{\nu}{k}\mathbf{B}^k = \mathbf{I} - \mathbf{B}^{\nu+1} + \sum_{k=1}^{\nu}\left(\binom{\nu}{k} - \binom{\nu}{k-1}\right)\mathbf{B}^k.$$

This yields

$$\|(\mathbf{I} - \mathbf{B})(\mathbf{I} + \mathbf{B})^\nu\| \leq 2 + \sum_{k=1}^{\nu}\left|\binom{\nu}{k} - \binom{\nu}{k-1}\right|.$$

Using $\binom{\nu}{k} \geq \binom{\nu}{k-1} \Leftrightarrow k \leq \frac{1}{2}(\nu+1)$ and $\binom{\nu}{k} \geq \binom{\nu}{\nu-k}$ we get (with $[\cdot]$ the round down operator):

$$\begin{aligned}
&\sum_{k=1}^{\nu} |\binom{\nu}{k} - \binom{\nu}{k-1}| \\
&= \sum_{1}^{[\frac{1}{2}(\nu+1)]} (\binom{\nu}{k} - \binom{\nu}{k-1}) + \sum_{[\frac{1}{2}(\nu+1)]+1}^{\nu} (\binom{\nu}{k-1} - \binom{\nu}{k}) \\
&= \sum_{1}^{[\frac{1}{2}\nu]} (\binom{\nu}{k} - \binom{\nu}{k-1}) + \sum_{m=1}^{[\frac{1}{2}\nu]} (\binom{\nu}{m} - \binom{\nu}{m-1}) \\
&= 2\sum_{k=1}^{[\frac{1}{2}\nu]} (\binom{\nu}{k} - \binom{\nu}{k-1}) = 2(\binom{\nu}{[\frac{1}{2}\nu]} - \binom{\nu}{0}).
\end{aligned}$$

An elementary analysis yields (cf., for example, Reusken[19])

$$\binom{\nu}{[\frac{1}{2}\nu]} \leq 2^{\nu}\sqrt{\frac{2}{\pi\nu}} \quad \text{for } \nu \geq 1.$$

Thus we have proved the bound. ■

Corollary 7.13 *Let $\|\cdot\|$ be any induced matrix norm. Assume that for a linear iterative method with iteration matrix $\mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell$ we have*

$$\|\mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell\| \leq 1 \tag{60}$$

Then for $\mathbf{S}_\ell := \mathbf{I} - \frac{1}{2}\mathbf{M}_\ell^{-1}\mathbf{A}_\ell$ the following smoothing property holds:

$$\|\mathbf{A}_\ell\mathbf{S}_\ell^\nu\| \leq 2\sqrt{\frac{2}{\pi\nu}}\,\|\mathbf{M}_\ell\|\,, \quad \nu = 1, 2, \ldots$$

Proof: Define $\mathbf{B} = \mathbf{I} - \mathbf{M}_\ell^{-1}\mathbf{A}_\ell$ and apply lemma 7.12:

$$\|\mathbf{A}_\ell\mathbf{S}_\ell^\nu\| \leq \|\mathbf{M}_\ell\|(\frac{1}{2})^\nu\|(\mathbf{I}-\mathbf{B})(\mathbf{I}+\mathbf{B})^\nu\| \leq 2\sqrt{\frac{2}{\pi\nu}}\|\mathbf{M}_\ell\|.$$

■

Remark 7.14 Note that in the smoother in corollary 7.13 we use damping with a factor $\frac{1}{2}$. Generalizations of the results in lemma 7.12 and corollary 7.13 are given in Nevanlinna[20], Hackbusch[21], Zulehner[22]. In Nevanlinna[20], Zulehner[22] it is shown that the damping factor $\frac{1}{2}$ can be replaced by an arbitrary damping factor $\omega \in (0, 1)$. Also note that in the smoothing property in corollary 7.13 we have a ν-dependence of the form $\nu^{-\frac{1}{2}}$, whereas in the symmetric case this is of the form ν^{-1}. It Hackbusch[21] it is shown that this loss of a factor $\nu^{\frac{1}{2}}$ when going to the nonsymmetric case is due to the fact that complex eigenvalues may occur. □

To verify the condition in (60) we will use the following elementary result:

Lemma 7.15 *If $\mathbf{E} \in \mathbb{R}^{m\times m}$ is such that there exists a $c > 0$ with*

$$\|\mathbf{E}\mathbf{x}\|_2^2 \le c\langle \mathbf{E}\mathbf{x}, \mathbf{x}\rangle \quad \textit{for all } \mathbf{x} \in \mathbb{R}^m$$

then we have $\|\mathbf{I} - \omega\mathbf{E}\|_2 \le 1$ for all $\omega \in [0, \frac{2}{c}]$.
Proof: Follows from:

$$\begin{aligned}\|(\mathbf{I} - \omega\mathbf{E})\mathbf{x}\|_2^2 &= \|\mathbf{x}\|_2^2 - 2\omega\langle \mathbf{E}\mathbf{x}, \mathbf{x}\rangle + \omega^2\|\mathbf{E}\mathbf{x}\|_2^2 \\ &\le \|\mathbf{x}\|_2^2 - \omega(\frac{2}{c} - \omega)\|\mathbf{E}\mathbf{x}\|_2^2 \\ &\le \|\mathbf{x}\|_2^2 \quad \text{if } \omega(\frac{2}{c} - \omega) \ge 0.\end{aligned}$$

■

We now use these results to derive a smoothing property for the Richardson method.
Theorem 7.16 *Assume that the bilinear form satisfies the usual conditions* (22)*. Let $\mathbf{A}_\ell$ be the stiffness matrix in* (26)*. There exist constants $\omega > 0$ and c independent of ℓ such that the following smoothing property holds:*

$$\|\mathbf{A}_\ell(\mathbf{I} - \omega h_\ell^{2-d}\mathbf{A}_\ell)^\nu\|_2 \le \frac{c}{\sqrt{\nu}}\|\mathbf{A}_\ell\|_2\,, \quad \nu = 1, 2, \ldots.$$

Proof: Using lemma 7.3, the inverse inequality and the ellipticity of the bilinear form we get, for arbitrary $\mathbf{x} \in \mathbb{R}^{n_\ell}$:

$$\begin{aligned}\|\mathbf{A}_\ell\mathbf{x}\|_2 = \max_{\mathbf{y}\in\mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell\mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{y}\|_2} &\le c\, h_\ell^{\frac{1}{2}d} \max_{v_\ell\in V_\ell} \frac{k(P_\ell\mathbf{x}, v_\ell)}{\|v_\ell\|_{L^2}} \\ &\le c\, h_\ell^{\frac{1}{2}d} \max_{v_\ell\in V_\ell} \frac{|P_\ell\mathbf{x}|_1|v_\ell|_1}{\|v_\ell\|_{L^2}} \le c\, h_\ell^{\frac{1}{2}d-1}|P_\ell\mathbf{x}|_1 \\ &\le c\, h_\ell^{\frac{1}{2}d-1}k(P_\ell\mathbf{x}, P_\ell\mathbf{x})^{\frac{1}{2}} = c\, h_\ell^{\frac{1}{2}d-1}\langle \mathbf{A}_\ell\mathbf{x}, \mathbf{x}\rangle^{\frac{1}{2}}.\end{aligned}$$

From this and lemma 7.15 it follows that there exists a constant $\omega > 0$ such that

$$\|\mathbf{I} - 2\omega h_\ell^{2-d}\mathbf{A}_\ell\|_2 \le 1 \quad \text{for all } \ell. \tag{61}$$

Define $\mathbf{M}_\ell := \frac{1}{2\omega}h_\ell^{d-2}\mathbf{I}$. From lemma 7.4 it follows that there exists a constant c_M independent of ℓ such that $\|\mathbf{M}_\ell\|_2 \le c_M\|\mathbf{A}_\ell\|_2$. Application of corollary 7.13 proves the result of the lemma. ■

We now consider the damped Jacobi method.
Theorem 7.17 *Assume that the bilinear form satisfies the usual conditions* (22)*. Let $\mathbf{A}_\ell$ be the stiffness matrix in* (26) *and $\mathbf{D}_\ell = \operatorname{diag}(\mathbf{A}_\ell)$. There exist constants $\omega > 0$ and c independent of ℓ such that the following smoothing property holds:*

$$\|\mathbf{A}_\ell(\mathbf{I} - \omega\mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_2 \le \frac{c}{\sqrt{\nu}}\|\mathbf{A}_\ell\|_2\,, \quad \nu = 1, 2, \ldots$$

Proof: We use the matrix norm induced by the vector norm $\|\mathbf{y}\|_D := \|\mathbf{D}_\ell^{\frac{1}{2}}\mathbf{y}\|_2$ for $\mathbf{y} \in \mathbb{R}^{n_\ell}$. Note that for $\mathbf{B} \in \mathbb{R}^{n_\ell\times n_\ell}$ we have $\|\mathbf{B}\|_D = \|\mathbf{D}_\ell^{\frac{1}{2}}\mathbf{B}\mathbf{D}_\ell^{-\frac{1}{2}}\|_2$. The inequalities

$$\|\mathbf{D}_\ell^{-1}\|_2 \le c_1\, h_\ell^{2-d}, \quad \kappa(\mathbf{D}_\ell) \le c_2 \tag{62}$$

hold with constants c_1, c_2 independent of ℓ. Using this in combination with lemma 7.3, the inverse inequality and the ellipticity of the bilinear form we get, for arbitrary $\mathbf{x} \in \mathbb{R}^{n_\ell}$:

$$\begin{aligned}
\|\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}\|_2 &= \max_{\mathbf{y}\in\mathbb{R}^{n_\ell}} \frac{\langle \mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}, \mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{y}\rangle}{\|\mathbf{y}\|_2} = \max_{\mathbf{y}\in\mathbb{R}^{n_\ell}} \frac{k(P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}, P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{y})}{\|\mathbf{y}\|_2} \\
&\le c\, h_\ell^{-1} \max_{\mathbf{y}\in\mathbb{R}^{n_\ell}} \frac{|P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}|_1 \|P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{y}\|_{L^2}}{\|\mathbf{y}\|_2} \\
&\le c\, h_\ell^{\frac{1}{2}d-1} |P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}|_1 \|\mathbf{D}_\ell^{-\frac{1}{2}}\|_2 \le c\, |P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}|_1 \\
&\le c\, k(P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}, P_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x})^{\frac{1}{2}} = c\, \langle \mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{x}, \mathbf{x}\rangle^{\frac{1}{2}}.
\end{aligned}$$

From this and lemma 7.15 it follows that there exists a constant $\omega > 0$ such that

$$\|\mathbf{I} - 2\omega\mathbf{D}_\ell^{-1}\mathbf{A}_\ell\|_D = \|\mathbf{I} - 2\omega\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}\|_2 \le 1 \quad \text{for all } \ell.$$

Define $\mathbf{M}_\ell := \frac{1}{2\omega}\mathbf{D}_\ell$. Application of corollary 7.13 with $\|\cdot\| = \|\cdot\|_D$ in combination with (62) yields

$$\begin{aligned}
\|\mathbf{A}_\ell(\mathbf{I} - \omega h_\ell \mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_2 &\le \kappa(\mathbf{D}_\ell^{\frac{1}{2}})\, \|\mathbf{A}_\ell(\mathbf{I} - \frac{1}{2}\mathbf{M}_\ell^{-1}\mathbf{A}_\ell)^\nu\|_D \\
&\le \frac{c}{\sqrt{\nu}}\|\mathbf{M}_\ell\|_D = \frac{c}{2\omega\sqrt{\nu}}\|\mathbf{D}_\ell\|_2 \le \frac{c}{\sqrt{\nu}}\|\mathbf{A}_\ell\|_2
\end{aligned}$$

and thus the result is proved. ∎

7.4 Multigrid contraction number

In this section we prove a bound for the contraction number in the Euclidean norm of the multigrid algorithm (31) with $\tau \ge 2$. We follow the analysis introduced by Hackbusch[1, 14]. Apart from the approximation and smoothing property that have been proved in the sections 7.2 and 7.3 we also need the following stability bound for the iteration matrix of the smoother:

$$\exists\, C_S \,:\, \|\mathbf{S}_\ell^\nu\|_2 \le C_S \quad \text{for all } \ell \text{ and } \nu. \tag{63}$$

Lemma 7.18 *Consider the Richardson method as in theorem 7.8 or theorem 7.16. In both cases* (63) *holds with* $C_S = 1$.
Proof: In the symmetric case (theorem 7.8) we have

$$\|\mathbf{S}_\ell\|_2 = \|\mathbf{I} - \frac{c_0}{\rho(\mathbf{A}_\ell)}\mathbf{A}_\ell\|_2 = \max_{\lambda\in\sigma(\mathbf{A}_\ell)} \big|1 - c_0\, \frac{\lambda}{\rho(\mathbf{A}_\ell)}\big| \le 1.$$

For the general case (theorem 7.16) we have, using (61):

$$\begin{aligned}
\|\mathbf{S}_\ell\|_2 &= \|\mathbf{I} - \omega h_\ell^{2-d}\mathbf{A}_\ell\|_2 = \|\frac{1}{2}\mathbf{I} + \frac{1}{2}(\mathbf{I} - 2\omega h_\ell^{2-d}\mathbf{A}_\ell)\|_2 \\
&\le \frac{1}{2} + \frac{1}{2}\|\mathbf{I} - 2\omega h_\ell^{2-d}\mathbf{A}_\ell\|_2 \le 1.
\end{aligned}$$

∎

Lemma 7.19 *Consider the damped Jacobi method as in theorem 7.9 or theorem 7.17. In both cases* (63) *holds.*
Proof: Both in the symmetric and nonsymmetric case we have

$$\|\mathbf{S}_\ell\|_D = \|\mathbf{D}_\ell^{\frac{1}{2}}(\mathbf{I} - \omega\mathbf{D}_\ell^{-1}\mathbf{A}_\ell)\mathbf{D}_\ell^{-\frac{1}{2}}\|_2 \leq 1$$

and thus

$$\|\mathbf{S}_\ell^\nu\|_2 \leq \|\mathbf{D}_\ell^{-\frac{1}{2}}(\mathbf{D}_\ell^{\frac{1}{2}}\mathbf{S}_\ell\mathbf{D}_\ell^{-\frac{1}{2}})^\nu\mathbf{D}_\ell^{\frac{1}{2}}\|_2 \leq \kappa(\mathbf{D}_\ell^{\frac{1}{2}})\,\|\mathbf{S}_\ell\|_D^\nu \leq \kappa(\mathbf{D}_\ell^{\frac{1}{2}})$$

Now note that $\mathbf{D}_\ell$ is uniformly (w.r.t. ℓ) well-conditioned. ∎

Using lemma 7.3 it follows that for $\mathbf{p}_\ell = P_\ell^{-1}P_{\ell-1}$ we have

$$C_{p,1}\|\mathbf{x}\|_2 \leq \|\mathbf{p}_\ell\mathbf{x}\|_2 \leq C_{p,2}\|\mathbf{x}\|_2 \quad \text{for all } \mathbf{x} \in \mathbb{R}^{n_{\ell-1}}. \tag{64}$$

with constants $C_{p,1} > 0$ and $C_{p,2}$ independent of ℓ.
We now formulate a main convergence result for the multigrid method.

Theorem 7.20 *Consider the multigrid method with iteration matrix given in* (49) *and parameter values $\nu_2 = 0$, $\nu_1 = \nu > 0$, $\tau \geq 2$. Assume that there are constants C_A, C_S and a monotonically decreasing function $g(\nu)$ with $g(\nu) \to 0$ for $\nu \to \infty$ such that for all ℓ:*

$$\|\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\|_2 \leq C_A\|\mathbf{A}_\ell\|_2^{-1} \tag{65a}$$

$$\|\mathbf{A}_\ell\mathbf{S}_\ell^\nu\|_2 \leq g(\nu)\,\|\mathbf{A}_\ell\|_2\,, \quad \nu \geq 1 \tag{65b}$$

$$\|\mathbf{S}_\ell^\nu\|_2 \leq C_S\,, \quad \nu \geq 1. \tag{65c}$$

For any $\xi^* \in (0,1)$ there exists a ν^* such that for all $\nu \geq \nu^*$

$$\|\mathbf{C}_{MG,\ell}\|_2 \leq \xi^*\,, \quad \ell = 0, 1, \ldots$$

holds.

Proof: For the two-grid iteration matrix we have

$$\|\mathbf{C}_{TG,\ell}\|_2 \leq \|\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\|_2\|\mathbf{A}_\ell\mathbf{S}_\ell^\nu\|_2 \leq C_A g(\nu).$$

Define $\xi_\ell = \|\mathbf{C}_{MG.\ell}\|_2$. From (49) we obtain $\xi_0 = 0$ and for $\ell \geq 1$:

$$\begin{aligned}
\xi_\ell &\leq C_A g(\nu) + \|\mathbf{p}_\ell\|_2\xi_{\ell-1}^\tau\|\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{S}_\ell^\nu\|_2 \\
&\leq C_A g(\nu) + C_{p,2}C_{p,1}^{-1}\xi_{\ell-1}^\tau\|\mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{S}_\ell^\nu\|_2 \\
&\leq C_A g(\nu) + C_{p,2}C_{p,1}^{-1}\xi_{\ell-1}^\tau\big(\|(\mathbf{I} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell)\mathbf{S}_\ell^\nu\|_2 + \|\mathbf{S}_\ell^\nu\|_2\big) \\
&\leq C_A g(\nu) + C_{p,2}C_{p,1}^{-1}\xi_{\ell-1}^\tau\big(C_A g(\nu) + C_S\big) \leq C_A g(\nu) + C^*\xi_{\ell-1}^\tau
\end{aligned}$$

with $C^* := C_{p,2}C_{p,1}^{-1}(C_A g(1) + C_S)$. Elementary analysis shows that for $\tau \geq 2$ and any $\xi^* \in (0,1)$ the sequence $x_0 = 0,\ x_i = C_A g(\nu) + C^* x_{i-1}^\tau,\ i \geq 1$, is bounded by ξ^* for $g(\nu)$ sufficiently small. ∎

Remark 7.21 Consider $\mathbf{A}_\ell,\ \mathbf{p}_\ell,\ \mathbf{r}_\ell$ as defined in (26), (29),(30). Assume that the variational problem (23) is such that the usual conditions (22) are satisfied. Moreover, the problem (23) and the corresponding dual problem are assumed to be H^2-regular. In the

multigrid method we use the Richardson or the damped Jacobi method described in section 7.3. *Then the assumptions* (65) *are fulfilled and thus for* $\nu_2 = 0$ *and* ν_1 *sufficiently large the multigrid W-cycle has a contraction number smaller than one independent of* ℓ. □

Remark 7.22 Let $\mathbf{C}_{MG,\ell}(\nu_2, \nu_1)$ be the iteration matrix of the multigrid method with ν_1 pre- and ν_2 postsmoothing iterations. With $\nu := \nu_1 + \nu_2$ we have

$$\rho\big(\mathbf{C}_{MG,\ell}(\nu_2, \nu_1)\big) = \rho\big(\mathbf{C}_{MG,\ell}(0, \nu)\big) \leq \|\mathbf{C}_{MG,\ell}(0, \nu)\|_2$$

Using theorem 7.20 we thus get, for $\tau \geq 2$, a bound for the *spectral radius* of the iteration matrix $\mathbf{C}_{MG,\ell}(\nu_2, \nu_1)$. □

Remark 7.23 The multigrid convergence analysis presented above assumes sufficient regularity (namely H^2-regularity) of the elliptic boundary value problem. There have been developed convergence analyses in which this regularity assumption is avoided and an h-independent convergence rate of multigrid is proved. These analyses are based on so-called subspace decomposition techniques. Two review papers on multigrid convergence proofs are Yserentant[23] and Xu[24]. □

7.5 Convergence analysis for symmetric positive definite problems

In this section we analyze the convergence of the multigrid method for the symmetric positive definite case, i.e., the stiffness matrix $\mathbf{A}_\ell$ is assumed to be symmetric positive definite. This property allows a refined analysis which proves that the contraction number of the multigrid method with $\tau \geq 1$ (the V-cycle is included !) and $\nu_1 = \nu_2 \geq 1$ pre- and postsmoothing iterations is bounded by a constant smaller than one independent of ℓ. The basic idea of this analysis is due to Braess[25] and is further simplified by Hackbusch[1, 14].

Throughout this section we make the following

Assumption 7.24 *In the bilinear form* $k(\cdot, \cdot)$ *in* (23) *we have* $\mathbf{b} = 0$ *and the conditions* (22) *are satisfied.*

Due to this the stiffness matrix $\mathbf{A}_\ell$ is symmetric positive definite and we can define the energy scalar product and corresponding norm:

$$\langle \mathbf{x}, \mathbf{y} \rangle_A := \langle \mathbf{A}_\ell \mathbf{x}, \mathbf{y} \rangle \,, \quad \|\mathbf{x}\|_A := \langle \mathbf{x}, \mathbf{x} \rangle_A^{\frac{1}{2}} \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^{n_\ell}.$$

We only consider smoothers with an iteration matrix $\mathbf{S}_\ell = \mathbf{I} - \mathbf{M}_\ell^{-1} \mathbf{A}_\ell$ in which $\mathbf{M}_\ell$ is symmetric positive definite. Important examples are the smoothers analyzed in section 7.3:

$$\text{Richardson method}: \ \mathbf{M}_\ell = c_0^{-1} \rho(\mathbf{A}_\ell) \mathbf{I} \,, \quad c_0 \in (0, 1] \tag{66a}$$

$$\text{Damped Jacobi}: \ \mathbf{M}_\ell = \omega^{-1} \mathbf{D}_\ell, \quad \omega \ \text{ as in thm. 7.9} \tag{66b}$$

$$\text{Symm. Gauss-Seidel}: \ \mathbf{M}_\ell = (\mathbf{D}_\ell - \mathbf{L}_\ell) \mathbf{D}_\ell^{-1} (\mathbf{D}_\ell - \mathbf{L}_\ell^T). \tag{66c}$$

For symmetric matrices $\mathbf{B}, \mathbf{C} \in \mathbb{R}^{m \times m}$ we use the notation $\mathbf{B} \leq \mathbf{C}$ iff $\langle \mathbf{B}\mathbf{x}, \mathbf{x} \rangle \leq \langle \mathbf{C}\mathbf{x}, \mathbf{x} \rangle$ for all $\mathbf{x} \in \mathbb{R}^m$.

Lemma 7.25 *For* $\mathbf{M}_\ell$ *as in* (66) *the following properties hold:*

$$\mathbf{A}_\ell \leq \mathbf{M}_\ell \quad \textit{for all } \ell \tag{67a}$$

$$\exists\, C_M: \quad \|\mathbf{M}_\ell\|_2 \leq C_M \|\mathbf{A}_\ell\|_2 \quad \textit{for all } \ell. \tag{67b}$$

Proof: For the Richardson method the result is trivial. For the damped Jacobi method we have $\omega \in (0, \rho(\mathbf{D}_\ell^{-1}\mathbf{A}_\ell)^{-1}]$ and thus $\omega\rho(\mathbf{D}_\ell^{-\frac{1}{2}}\mathbf{A}_\ell\mathbf{D}_\ell^{-\frac{1}{2}}) \leq 1$. This yields $\mathbf{A}_\ell \leq \omega^{-1}\mathbf{D}_\ell = \mathbf{M}_\ell$. The result in (67b) follows from $\|\mathbf{D}_\ell\|_2 \leq \|\mathbf{A}_\ell\|_2$. For the symmetric Gauss-Seidel method the results (67a) follows from $\mathbf{M}_\ell = \mathbf{A}_\ell + \mathbf{L}_\ell\mathbf{D}_\ell^{-1}\mathbf{L}_\ell^T$ and the result in (67b) is proved in (59). ■

We introduce the following *modified approximation property*:

$$\exists\, \tilde{C}_A: \quad \|\mathbf{M}_\ell^{\frac{1}{2}}\big(\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\big)\mathbf{M}_\ell^{\frac{1}{2}}\|_2 \leq \tilde{C}_A \quad \text{for } \ell = 1, 2, \ldots \tag{68}$$

We note that the standard approximation property (54) implies the result (68) if we consider the smoothers in (66):

Lemma 7.26 *Consider* $\mathbf{M}_\ell$ *as in* (66) *and assume that the approximation property* (54) *holds. Then* (68) *holds with* $\tilde{C}_A = C_M C_A$.

Proof: Trivial. ■

One easily verifies that for the smoothers in (66) the modified approximation property (68) implies the standard approximation property (54) if $\kappa(\mathbf{M}_\ell)$ is uniformly (w.r.t. ℓ) bounded. The latter property holds for the Richardson and the damped Jacobi method.

We will analyze the convergence of the two-grid and multigrid method using the energy scalar product. For matrices $\mathbf{B},\ \mathbf{C} \in \mathbb{R}^{n_\ell \times n_\ell}$ that are symmetric w.r.t. $\langle\cdot,\cdot\rangle_A$ we use the notation $\mathbf{B} \leq_A \mathbf{C}$ iff $\langle\mathbf{B}\mathbf{x},\mathbf{x}\rangle_A \leq \langle\mathbf{C}\mathbf{x},\mathbf{x}\rangle_A$ for all $\mathbf{x} \in \mathbb{R}^{n_\ell}$. Note that $\mathbf{B} \in \mathbb{R}^{n_\ell \times n_\ell}$ is symmetric w.r.t. $\langle\cdot,\cdot\rangle_A$ iff $(\mathbf{A}_\ell\mathbf{B})^T = \mathbf{A}_\ell\mathbf{B}$ holds. We also note the following elementary property for symmetric matrices $\mathbf{B},\ \mathbf{C} \in \mathbb{R}^{n_\ell \times n_\ell}$:

$$\mathbf{B} \leq \mathbf{C} \ \Leftrightarrow\ \mathbf{B}\mathbf{A}_\ell \leq_A \mathbf{C}\mathbf{A}_\ell. \tag{69}$$

We now turn to the two-grid method. For the coarse grid correction we introduce the notation $\mathbf{Q}_\ell := \mathbf{I} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell$. For symmetry reasons we only consider $\nu_1 = \nu_2 = \frac{1}{2}\nu$ with $\nu > 0$ even. The iteration matrix of the two-grid method is given by

$$\mathbf{C}_{TG,\ell} = \mathbf{C}_{TG,\ell}(\nu) = \mathbf{S}_\ell^{\frac{1}{2}\nu}\mathbf{Q}_\ell\mathbf{S}_\ell^{\frac{1}{2}\nu}.$$

Due the symmetric positive definite setting we have the following fundamental property:

Theorem 7.27 *The matrix* $\mathbf{Q}_\ell$ *is an orthogonal projection w.r.t.* $\langle\cdot,\cdot\rangle_A$.

Proof: Follows from

$$\mathbf{Q}_\ell^2 = \mathbf{Q}_\ell \quad \text{and} \quad (\mathbf{A}_\ell\mathbf{Q}_\ell)^T = \mathbf{A}_\ell\mathbf{Q}_\ell.$$

■

As an direct consequence we have

$$0 \leq_A \mathbf{Q}_\ell \leq_A \mathbf{I}. \tag{70}$$

The next lemma gives another characterization of the modified approximation property:

Lemma 7.28 *The property* (68) *is equivalent to*

$$0 \leq_A \mathbf{Q}_\ell \leq_A \tilde{C}_A\mathbf{M}_\ell^{-1}\mathbf{A}_\ell \quad \textit{for } \ell = 1, 2, \ldots \tag{71}$$

Proof: Using (69) we get

$$\begin{aligned}
&\|\mathbf{M}_\ell^{\frac{1}{2}}\big(\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\big)\mathbf{M}_\ell^{\frac{1}{2}}\|_2 \le \tilde{C}_A \quad \text{for all } \ell \\
\Leftrightarrow\ &-\tilde{C}_A\mathbf{I} \le \mathbf{M}_\ell^{\frac{1}{2}}\big(\mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\big)\mathbf{M}_\ell^{\frac{1}{2}} \le \tilde{C}_A\mathbf{I} \quad \text{for all } \ell \\
\Leftrightarrow\ &-\tilde{C}_A\mathbf{M}_\ell^{-1} \le \mathbf{A}_\ell^{-1} - \mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell \le \tilde{C}_A\mathbf{M}_\ell^{-1} \quad \text{for all } \ell \\
\Leftrightarrow\ &-\tilde{C}_A\mathbf{M}_\ell^{-1}\mathbf{A}_\ell \le_A \mathbf{Q}_\ell \le_A \tilde{C}_A\mathbf{M}_\ell^{-1}\mathbf{A}_\ell \quad \text{for all } \ell.
\end{aligned}$$

In combination with (70) this proves the result. ■

We now present a convergence result for the two-grid method:

Theorem 7.29 *Assume that* (67a) *and* (68) *hold. Then we have*

$$\begin{aligned}
\|\mathbf{C}_{TG,\ell}(\nu)\|_A &\le \max_{y\in[0,1]} y(1-\tilde{C}_A^{-1}y)^\nu \\
&= \begin{cases} (1-\tilde{C}_A^{-1})^\nu & \text{if } \ \nu \le \tilde{C}_A - 1 \\ \frac{\tilde{C}_A}{\nu+1}\big(\frac{\nu}{\nu+1}\big)^\nu & \text{if } \ \nu \ge \tilde{C}_A - 1. \end{cases}
\end{aligned} \tag{72}$$

Proof: Define $\mathbf{X}_\ell := \mathbf{M}_\ell^{-1}\mathbf{A}_\ell$. This matrix is symmetric w.r.t. the energy scalar product and from (67a) it follows that

$$0 \le_A \mathbf{X}_\ell \le_A \mathbf{I} \tag{73}$$

holds. From lemma 7.28 we obtain $0 \le_A \mathbf{Q}_\ell \le_A \tilde{C}_A\mathbf{X}_\ell$. Note that due to this, (73) and the fact that $\mathbf{Q}_\ell$ is an A-orthogonal projection which is not identically zero we get

$$\tilde{C}_A \ge 1. \tag{74}$$

Using (70) we get

$$0 \le_A \mathbf{Q}_\ell \le_A \alpha\tilde{C}_A\mathbf{X}_\ell + (1-\alpha)\mathbf{I} \quad \text{for all } \ \alpha \in [0,1]. \tag{75}$$

Hence, using $\mathbf{S}_\ell = \mathbf{I} - \mathbf{X}_\ell$ we have

$$0 \le_A \mathbf{C}_{TG,\ell}(\nu) \le_A (\mathbf{I}-\mathbf{X}_\ell)^{\frac{1}{2}\nu}\big(\alpha\tilde{C}_A\mathbf{X}_\ell + (1-\alpha)\mathbf{I}\big)(\mathbf{I}-\mathbf{X}_\ell)^{\frac{1}{2}\nu}$$

for all $\alpha \in [0,1]$, and thus

$$\|\mathbf{C}_{TG,\ell}(\nu)\|_A \le \min_{\alpha\in[0,1]}\max_{x\in[0,1]}\big(\alpha\tilde{C}_A x + (1-\alpha)\big)(1-x)^\nu.$$

A minimax result (cf. Sion[26]) implies that in the previous expression the min and max operations can be interchanged. A simple computation yields

$$\begin{aligned}
&\max_{x\in[0,1]}\min_{\alpha\in[0,1]}\big(\alpha\tilde{C}_A x + (1-\alpha)\big)(1-x)^\nu \\
&= \max\Big\{\max_{x\in[0,\tilde{C}_A^{-1}]}\tilde{C}_A x(1-x)^\nu\,,\ \max_{x\in[\tilde{C}_A^{-1},1]}(1-x)^\nu\Big\} \\
&= \max_{x\in[0,\tilde{C}_A^{-1}]}\tilde{C}_A x(1-x)^\nu = \max_{y\in[0,1]} y(1-\tilde{C}_A^{-1}y)^\nu.
\end{aligned}$$

This proves the inequality in (72). An elementary computation shows that the equality in (72) holds. ∎

We now show that the approach used in the convergence analysis of the two-grid method in theorem 7.29 can also be used for the multigrid method.
We start with an elementary result concerning a fixed point iteration that will be used in theorem 7.31.

Lemma 7.30 *For given constants $c > 1$, $\nu \geq 1$ define $g : [0,1) \to \mathbb{R}$ by*

$$g(\xi) = \begin{cases} (1 - \frac{1}{c})^{\nu} & \text{if } \ 0 \leq \xi < 1 - \frac{\nu}{c-1} \\ \frac{c}{\nu+1}\big(\frac{\nu}{\nu+1}\big)^{\nu}(1-\xi)\big(1 + \frac{1}{c}\frac{\xi}{1-\xi}\big)^{\nu+1} & \text{if } \ 1 - \frac{\nu}{c-1} \leq \xi < 1. \end{cases} \tag{76}$$

For $\tau \in \mathbb{N}$, $\tau \geq 1$, define the sequence $\xi_{\tau,0} = 0, \quad \xi_{\tau,i+1} = g(\xi_{\tau,i}^{\tau})$ for $i \geq 1$. The following holds:

- *$\xi \to g(\xi)$ is continuous and increasing on $[0,1)$.*
- *For $c = \tilde{C}_A$, $g(0)$ coincides with the upper bound in* (72)*.*
- *$g(\xi) = \xi$ iff $\xi = \dfrac{c}{c+\nu}$.*
- *The sequence $(\xi_{\tau,i})_{i\geq 0}$ is monotonically increasing, and $\xi_{\tau}^{*} := \lim\limits_{i\to\infty} \xi_{\tau,i} < 1$.*
- *$\big((\xi_{\tau}^{*})^{\tau}, \xi_{\tau}^{*}\big)$ is the first intersection point of the graphs of $g(\xi)$ and $\xi^{\frac{1}{\tau}}$.*
- *$\dfrac{c}{c+\nu} = \xi_1^{*} \geq \xi_2^{*} \geq \ldots \geq \xi_{\infty}^{*} = g(0)$.*

Proof: Elementary calculus. ∎

As an illustration for two pairs (c, ν) we show the graph of the function g in Fig. 9.

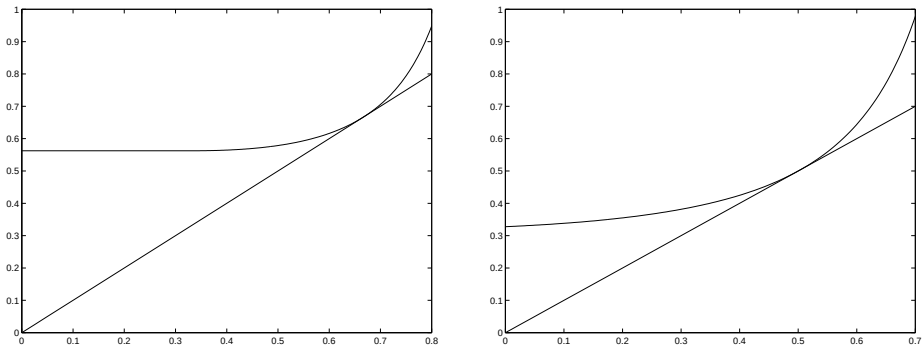


Figure 9. Function $g(\xi)$ for $\nu = 2,\ c = 4$ (left) and $\nu = 4,\ c = 4$ (right).

Theorem 7.31 *We take $\nu_1 = \nu_2 = \nu$ and consider the multigrid algorithm with iteration matrix $\mathbf{C}_{MG,\ell} = \mathbf{C}_{MG,\ell}(\nu,\tau)$ as in* (49)*. Assume that* (67a) *and* (68) *hold. For $c = \tilde{C}_A$, $\nu \geq 2$ and τ as in* (49) *let $\xi_\tau^* \leq \frac{c}{c+\nu}$ be the fixed point defined in lemma 7.30. Then*

$$\|\mathbf{C}_{MG,\ell}\|_A \leq \xi_\tau^*$$

holds.

Proof: From (49) we have

$$\begin{aligned}\mathbf{C}_{MG,\ell} &= \mathbf{S}_\ell^{\frac{1}{2}\nu}\big(\mathbf{I} - \mathbf{p}_\ell(\mathbf{I} - \mathbf{C}_{MG,\ell-1}^\tau)\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\big)\mathbf{S}_\ell^{\frac{1}{2}\nu} \\ &= \mathbf{S}_\ell^{\frac{1}{2}\nu}(\mathbf{Q}_\ell + \mathbf{R}_\ell)\mathbf{S}_\ell^{\frac{1}{2}\nu}\,, \quad \mathbf{R}_\ell := \mathbf{p}_\ell\mathbf{C}_{MG,\ell-1}^\tau\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell.\end{aligned}$$

The matrices $\mathbf{S}_\ell$ and $\mathbf{Q}_\ell$ are symmetric w.r.t. $\langle\cdot,\cdot\rangle_A$. If $\mathbf{C}_{MG,\ell-1}$ is symmetric w.r.t. $\langle\cdot,\cdot\rangle_{A_{\ell-1}}$ then from

$$(\mathbf{A}_\ell\mathbf{R}_\ell)^T = \big[(\mathbf{A}_\ell\mathbf{p}_\ell\mathbf{A}_{\ell-1}^{-1})(\mathbf{A}_{\ell-1}\mathbf{C}_{MG,\ell-1}^\tau)(\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell)\big]^T = \mathbf{A}_\ell\mathbf{R}_\ell$$

it follows that $\mathbf{R}_\ell$ is symmetric w.r.t. $\langle\cdot,\cdot\rangle_A$, too. By induction we conclude that for all ℓ the matrices $\mathbf{R}_\ell$ and $\mathbf{C}_{MG,\ell}$ are symmetric w.r.t. $\langle\cdot,\cdot\rangle_A$. Note that

$$0 \leq_A \mathbf{C}_{MG,\ell-1}^\tau \;\Leftrightarrow\; 0 \leq \mathbf{C}_{MG,\ell-1}^\tau\mathbf{A}_{\ell-1}^{-1} \;\Leftrightarrow\; 0 \leq \mathbf{p}_\ell\mathbf{C}_{MG,\ell-1}^\tau\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell \;\Leftrightarrow\; 0 \leq_A \mathbf{R}_\ell$$

holds. Thus, by induction and using $0 \leq_A \mathbf{Q}_\ell$ we get

$$0 \leq_A \mathbf{Q}_\ell + \mathbf{R}_\ell\,, \quad 0 \leq_A \mathbf{C}_{MG,\ell} \quad \text{for all } \ell. \tag{77}$$

For $\ell \geq 0$ define $\xi_\ell := \|\mathbf{C}_{MG,\ell}\|_A$. Hence, $0 \leq_A \mathbf{C}_{MG,\ell} \leq_A \xi_\ell\mathbf{I}$ holds. For arbitrary $\mathbf{x} \in \mathbb{R}^{n_\ell}$ we have

$$\begin{aligned}\langle\mathbf{R}_\ell\mathbf{x},\mathbf{x}\rangle_A &= \langle\mathbf{C}_{MG,\ell-1}^\tau\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{x}, \mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{x}\rangle_{A_{\ell-1}} \\ &\leq \xi_{\ell-1}^\tau\langle\mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{x}, \mathbf{A}_{\ell-1}^{-1}\mathbf{r}_\ell\mathbf{A}_\ell\mathbf{x}\rangle_{A_{\ell-1}} = \xi_{\ell-1}^\tau\langle\mathbf{x},(\mathbf{I}-\mathbf{Q}_\ell)\mathbf{x}\rangle_A\end{aligned}$$

and thus

$$\mathbf{R}_\ell \leq_A \xi_{\ell-1}^\tau(\mathbf{I} - \mathbf{Q}_\ell) \tag{78}$$

holds. Define $\mathbf{X}_\ell := \mathbf{M}_\ell^{-1}\mathbf{A}_\ell$. Using (75), (77) and (78) we get

$$\begin{aligned}0 \leq_A \mathbf{Q}_\ell + \mathbf{R}_\ell &\leq_A (1-\xi_{\ell-1}^\tau)\mathbf{Q}_\ell + \xi_{\ell-1}^\tau\mathbf{I} \\ &\leq_A (1-\xi_{\ell-1}^\tau)\big(\alpha\tilde{C}_A\mathbf{X}_\ell + (1-\alpha)\mathbf{I}\big) + \xi_{\ell-1}^\tau\mathbf{I} \quad \text{for all } \alpha \in [0,1].\end{aligned}$$

Hence, for all $\alpha \in [0,1]$ we have

$$0 \leq_A \mathbf{C}_{MG,\ell} \leq_A (\mathbf{I}-\mathbf{X}_\ell)^{\frac{1}{2}\nu}\big[(1-\xi_{\ell-1}^\tau)\big(\alpha\tilde{C}_A\mathbf{X}_\ell + (1-\alpha)\mathbf{I}\big) + \xi_{\ell-1}^\tau\mathbf{I}\big](\mathbf{I}-\mathbf{X}_\ell)^{\frac{1}{2}\nu}.$$

This yields

$$\xi_\ell \leq \min_{\alpha\in[0,1]}\max_{x\in[0,1]}\big[(1-\xi_{\ell-1}^\tau)\big(\alpha\tilde{C}_A x + 1 - \alpha\big) + \xi_{\ell-1}^\tau\big](1-x)^\nu.$$

As in the proof of theorem 7.29 we can interchange the min and max operations in the previous expression. A simple computation shows that for $\xi \in [0,1]$ we have

$$\begin{aligned}&\max_{x\in[0,1]}\min_{\alpha\in[0,1]}\big[(1-\xi)\big(\alpha\tilde{C}_A x + 1 - \alpha\big) + \xi\big](1-x)^\nu \\ &= \max\Big\{\max_{x\in[0,\tilde{C}_A^{-1}]}\big((1-\xi)\tilde{C}_A x + \xi\big)(1-x)^\nu\,,\ \max_{x\in[\tilde{C}_A^{-1},1]}(1-x)^\nu\Big\} = g(\xi)\end{aligned}$$

where $g(\xi)$ is the function defined in lemma 7.30 with $c = \tilde{C}_A$. Thus ξ_ℓ satisfies $\xi_0 = 0$ and $\xi_\ell \leq g(\xi_{\ell-1}^\tau)$ for $\ell \geq 1$. Application of the results in lemma 7.30 completes the proof. ■

The bound ξ_τ^* for the multigrid contraction number in theorem 7.31 decreases if τ increases. Moreover, for $\tau \to \infty$ the bound converges to the bound for the two-grid contraction number in theorem 7.29.

Corollary 7.32 *Consider $\mathbf{A}_\ell,\ \mathbf{p}_\ell,\ \mathbf{r}_\ell$ as defined in* (26), (29),(30)*. Assume that the variational problem* (23) *is such that $\mathbf{b} = 0$ and that the usual conditions* (22) *are satisfied. Moreover, the problem is assumed to be H^2-regular. In the multigrid method we use one of the smoothers* (66)*. Then the assumptions* (67a) *and* (68) *are satisfied and thus for $\nu_1 = \nu_2 \geq 1$ the multigrid V-cycle has a contraction number (w.r.t. $\|\cdot\|_A$) smaller than one independent of ℓ.* □

8 Convergence Analysis for Stokes Problems

The multigrid method for the Stokes problem can be analyzed along the same lines as in section 7.4, i.e., based on a smoothing and approximation property. For the Stokes problem an analysis which proves convergence of the V-cycle is *not* known. In other words, results as presented for scalar elliptic problems in section 7.5 are not known for the Stokes equation.

We briefly outline the convergence results available for multigrid applied to the Stokes problem. For a detailed treatment we refer to the literature, for example to Verfürth[27], Larin[28], Zulehner[11]. As in section 7 we assume that the family of triangulations $\{\mathcal{T}_{h_\ell}\}$ is quasi-uniform and that $h_{\ell-1}/h_\ell$ is uniformly bounded w.r.t. ℓ. We assume H^2-regularity of the Stokes problem, i.e., for the solution $(\vec{u}, p)$ of (38) we have

$$\|\vec{u}\|_{H^2} + \|p\|_{H^1} \leq c\|\vec{f}\|_{L^2}$$

with a constant c independent of $\vec{f} \in L^2(\Omega)^d$. The finite element spaces $\mathbf{V}_\ell,\ Q_\ell$ should have the approximation property

$$\inf_{\vec{v}\in\mathbf{V}_\ell} \|\vec{u} - \vec{v}\|_{H^1} + \inf_{q\in Q_\ell} \|p - q\|_{L^2} \leq c\, h_\ell \big(\|\vec{u}\|_{H^2} + \|p\|_{H^1}\big),$$

for all $\vec{u} \in (H^2(\Omega) \cap H_0^1(\Omega))^d,\ p \in H^1(\Omega) \cap L_0^2(\Omega)$. This holds, for example, for the Hood-Taylor pair of finite element spaces. Let $\mathcal{A}_\ell$ be the Stokes stiffness matrix as in (40) and $\mathcal{S}_\ell$ the iteration matrix of the smoother. The prolongation P_ℓ is as in (41). For the restriction R_ℓ we take the adjoint of the prolongation. The iteration matrix of the two-grid method with $\nu = \nu_1$ pre-smoothing and $\nu_2 = 0$ post-smoothing iterations is given by

$$\mathcal{M}_\ell = (I - P_\ell \mathcal{A}_{\ell-1}^{-1} R_\ell \mathcal{A}_\ell)\mathcal{S}_\ell^\nu.$$

For the analysis we have to introduce a suitable scaled Euclidean norm defined by

$$\left\| \begin{pmatrix} \mathbf{u}_\ell \\ \mathbf{p}_\ell \end{pmatrix} \right\|_h^2 := \|\mathbf{u}_\ell\|^2 + h_\ell^2 \|\mathbf{p}_\ell\|^2 = \left\| \Lambda_\ell \begin{pmatrix} \mathbf{u}_\ell \\ \mathbf{p}_\ell \end{pmatrix} \right\|^2 \quad \text{with } \Lambda_\ell := \begin{pmatrix} I_{n_\ell} & 0 \\ 0 & h_\ell I_{m_\ell} \end{pmatrix}. \tag{79}$$

Furthermore we introduce the scaled matrices

$$\tilde{\mathcal{A}}_\ell := \Lambda_\ell^{-1} \mathcal{A}_\ell \Lambda_\ell^{-1} = \begin{pmatrix} A_\ell & h_\ell^{-1} B_\ell^T \\ h_\ell^{-1} B_\ell & 0 \end{pmatrix}, \quad \tilde{\mathcal{S}}_\ell := \Lambda_\ell \mathcal{S}_\ell \Lambda_\ell^{-1}.$$

Using these definitions we obtain

$$\begin{aligned}\|\mathcal{M}_\ell\|_h &= \|\Lambda_\ell(\mathcal{A}_\ell^{-1} - P_\ell\mathcal{A}_{\ell-1}^{-1}R_\ell)\Lambda_\ell\Lambda_\ell^{-1}\mathcal{A}_\ell\mathcal{S}_\ell^\nu\Lambda_\ell^{-1}\| \\ &\le \|\Lambda_\ell(\mathcal{A}_\ell^{-1} - P_\ell\mathcal{A}_{\ell-1}^{-1}R_\ell)\Lambda_\ell\|\,\|\tilde{\mathcal{A}}_\ell\tilde{\mathcal{S}}_\ell^\nu\|.\end{aligned}$$

In Larin[28] the *approximation property*

$$\|\Lambda_\ell(\mathcal{A}_\ell^{-1} - P_\ell\mathcal{A}_{\ell-1}^{-1}R_\ell)\Lambda_\ell\| \le ch_\ell^2 \tag{80}$$

is proved. In that paper it is also shown (using an analysis along the same lines as in section 7.3) that for the Braess-Sarazin method in which the system in (45) is solved exactly, we have a smoothing property

$$\|\tilde{\mathcal{A}}_\ell\tilde{\mathcal{S}}_\ell^\nu\| \le \frac{ch_\ell^{-2}}{e(\nu-2)+1} \quad \text{for } \nu \ge 2. \tag{81}$$

In Zulehner[11] a smoothing property for the Braess-Sarazin method with an *inexact* (but sufficiently accurate) inner solve for the system (45) is proved:

$$\|\tilde{\mathcal{A}}_\ell\tilde{\mathcal{S}}_\ell^\nu\| \le \frac{ch_\ell^{-2}}{\nu-1} \quad \text{for } \nu \ge 2. \tag{82}$$

Combining the approximation property in (80) with the smoothing property (81) or (82) we obtain a bound for the contraction number of the two-grid iteration matrix:

$$\|\mathcal{M}_\ell\|_h \le \frac{c_A}{\nu-1} \quad \text{for } \nu \ge 2$$

with a constant c_A independent of ℓ and ν. Thus we have a two-grid convergence with a rate independent of ℓ if the number of smoothing iterations ν is sufficiently high. Using an analysis as in section 7.4 one can derive a convergence result for the multigrid W-cycle method.

A smoothing property of the form (81), (82) for the Vanka smoother is *not* known in the literature. A theoretical analysis which proves convergence of the multigrid method with a Vanka smoother for the Stokes equations is not available.

References

1. Wolfgang Hackbusch, *Multigrid Methods and Applications*, vol. 4 of *Springer Series in Computational Mathematics*, Springer, Berlin, Heidelberg, 1985.
2. P. Wesseling, *An introduction to Multigrid Methods*, Wiley, Chichester, 1992.
3. W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial (2nd ed.)*, SIAM, Philadelphia, 2000.
4. U. Trottenberg C. W. Oosterlee and A. Schüler, (Eds.), *Multigrid*, Academic Press, London, 2001.
5. J. H. Bramble, *Multigrid Methods*, Longman, Harlow, 1993.
6. Wolfgang Hackbusch, *Elliptic Differential Equations: Theory and Numerical Treatment*, vol. 18 of *Springer Series in Computational Mathematics*, Springer, Berlin, 1992.
7. Christian Großmann and H. G. Roos, *Numerik Partieller Differentialgleichungen*, Teubner, Stuttgart, 2. edition, 1994.

8. Jürgen Bey, *Tetrahedral Grid Refinement*, Computing, **55**, no. 4, 355–378, 1995.
9. Dietrich Braess, M. Dryja, and Wolfgang Hackbusch, *A multigrid method for nonconforming FE-discretisations with application to non-matching grids*, Computing, **63**, 1–25, 1999.
10. Dietrich Braess and R. Sarazin, *An efficient smoother for the Stokes problem*, Applied Numerical Mathematics, **23**, 3–19, 1997.
11. Walter Zulehner, *A class of smoothers for saddle point problems*, Computing, **65**, 227–246, 2000.
12. S. Vanka, *Block-implicit multigrid calculation of two-dimensional recirculating flows*, Computer Methods in Appl. Mech. and Eng., **59**, 29–48, 1986.
13. Volker John and G. Matthies, *Higher order finite element discretizations in a benchmark problem for incompressible flows*, Int. J. Num. Meth. Fluids, **37**, 885–903, 2001.
14. Wolfgang Hackbusch, *Iterative Solution of Large Sparse Systems of Equations*, vol. 95 of *Applied Mathematical Sciences*, Springer, New York, 1994.
15. Gabriel Wittum, *On the Robustness of ILU-Smoothing*, SIAM J. Sci. Stat. Comp., **10**, 699–717, 1989.
16. Gabriel Wittum, *Linear iterations as smoothers in multigrid methods : Theory with applications to incomplete decompositions*, Impact Comput. Sci. Eng., **1**, 180–215, 1989.
17. O. Bröker, M. Grote, C. Mayer, and A. Reusken, *Robust parallel smoothing for multigrid via sparse approximate inverses*, SIAM J. Sci. Comput., **32**, 1395–1416, 2001.
18. Arnold Reusken, *On maximum norm convergence of multigrid methods for two-point boundary value problems*, SIAM J. Numer. Anal., **29**, 1569–1578, 1992.
19. Arnold Reusken, *The smoothing property for regular splittings*, in: Incomplete Decompositions : (ILU)-Algorithms, Theory and Applications, W. Hackbusch and G. Wittum, (Eds.), vol. 41 of *Notes on Numerical Fluid Mechanics*, pp. 130–138, Vieweg, Braunschweig, 1993.
20. Olavi Nevanlinna, *Convergence of Iterations for Linear Equations*, Birkhäuser, Basel, 1993.
21. W. Hackbusch, *A note on Reusken's Lemma*, Computing, **55**, 181–189, 1995.
22. E. Ecker and W. Zulehner, *On the smoothing property of multi-grid methods in the non-symmetric case*, Numerical Linear Algebra with Applications, **3**, 161–172, 1996.
23. Harry Yserentant, *Old and new convergence proofs of multigrid methods*, Acta Numerica, pp. 285–326, 1993.
24. Jinchao Xu, *Iterative methods by space decomposition and subspace correction*, SIAM Review, **34**, 581–613, 1992.
25. Dietrich Braess and Wolfgang Hackbusch, *A New Convergence Proof for the Multigrid Method including the V-Cycle*, SIAM J. Numer. Anal., **20**, 967–975, 1983.
26. Maurice Sion, *On general minimax theorems*, Pacific J. of Math., **8**, 171–176, 1958.
27. Rüdiger Verfürth, *Error estimates for a mixed finite element approximation of Stokes problem*, RAIRO Anal. Numer., **18**, 175–182, 1984.
28. Maxim Larin and Arnold Reusken, *A comparative study of efficient iterative solvers for generalized Stokes equations*, Numer. Linear Algebra Appl., **15**, 13–34, 2008.

Wavelets and Their Application for the Solution of Poisson's and Schrödinger's Equation

Stefan Goedecker

Department of Physics
University of Basel, Switzerland
E-mail: stefan.goedecker@unibas.ch

Wavelets can be used as a basis set for the solution of partial differential equations. After introducing the theoretical framework of wavelet theory, we will show how they can be used to solve Poisson's equation and Schrödinger's equation in an efficient way.

1 Wavelets, an Optimal Basis Set

The preferred way to solve partial differential equations is to express the solution as a linear combination of so-called basis functions. These basis functions can for instance be plane waves, Gaussians or finite elements. Having discretized the differential equation in this way makes it amenable to a numerical solution. In the case of Poisson's equation one obtains for instance a linear system of equation, in the case of Schrödinger's equation one obtains an eigenvalue problem. This procedure is usually more stable than other methods which do not involve basis functions, such as finite difference methods. Wavelets[3,2] are just another basis set which however offers considerable advantages over alternative basis sets and allows us to attack problems not accessible with conventional numerical methods. Its main advantages are:

- The basis set can be improved in an systematic way:
  If one wants the solution of the differential equation with higher accuracy one can just add more wavelets in the expansion of the solution. This will not lead to any numerical instabilities as one encounters for instance with Gaussians. The accuracy of the solution is determined by one single parameter similar to the minimal wavelength determining the accuracy of a plane wave expansion. In the case of the Gaussian type basis sets used in quantum chemistry there are many parameters which determine the accuracy and it is frequently not obvious which one has the largest leverage to improve upon the accuracy.

- Different resolutions can be used in different regions of space:
  If the solution of the differential equation is varying particularly rapidly in a particular region of space one can increase the resolution in this region by adding more high resolution wavelets centered around this region. This varying resolution is for instance not possible with plane waves, which give the same resolution in the whole computational volume.

- The coupling between different resolution levels is easy:
  Finite elements can also be used with varying resolution levels. The resulting highly structured grids lead however to very complicated matrix structures, requiring indirect

indexing of most arrays. In the case of wavelets, in contrast, the coupling between different resolution levels is rather easy.

- There are few topological constraints for increased resolution regions:
  The regions of increased resolution can be chosen arbitrarily, the only requirement being that a region of higher resolution be contained in a region of the next lower resolution. If one uses for instance generalized plane waves in connection with curvilinear coordinates[1] to obtain varying resolution one has the requirement that the varying resolution grid can be obtained by a mapping from a equally spaced grid. Increasing the resolution in one region requires decreasing the resolution in some other region.

- The Laplace operator is diagonally dominant in an appropriate wavelet basis:
  This allows for a simple but efficient preconditioning scheme for equations such as Poisson's or Schrödinger's equation which contains the Laplacian as well. As a result the number of iterations needed in the iterative solution of the linear algebra equations corresponding to these differential equations is fairly small and independent of the maximal resolution. No such easy and efficient preconditioning scheme is known for other varying resolution schemes such as finite elements, Gaussians or generalized plane waves with curvilinear coordinates.

- The matrix elements of the Laplace operator are very easy to calculate:
  The requirement that the matrix elements can easily be calculated is essential for any basis set and therefore fulfilled by all standard basis sets. For the case of wavelets it is however particularly easy since they can be calculated on the fly by simple scaling arguments and therefore need not be stored in memory.

- The numerical effort scales linearly with respect to system size:
  Three-dimensional problems of realistic size require usually a very large number of basis functions. It is therefore of utmost importance, that the numerical effort scales only linearly (and not quadratically or cubically) with respect to the number of basis functions. If one uses iterative matrix techniques, this linear scaling can only be obtained if two requirements are fulfilled, namely that the matrix vector multiplications which are necessary for all iterative methods can be done with linear scaling and that the number of matrix vector multiplications is independent of the problem size. The first requirement is fulfilled if either the matrix representing the differential operator is sparse or can be transformed into sparse form by a transformation which has linear scaling, a requirement fulfilled by wavelets. The second requirement is related to the availability of a good preconditioning scheme. Since such a scheme exists, the conditioning number of the involved matrices do not vary strongly with respect to the problem size and the number of iterations (i.e. matrix vector multiplications) is independent of the problem size.

2 A First Tour of Some Wavelet Families

Many families of wavelets have been proposed in the mathematical literature. If one wants to use wavelets for the solution of differential equations, one therefore has to choose one specific family which is most advantageous for the intended application. Within one family

there are also members of different degree. Without going into any detail at this point we will in the following just show some plots of some common wavelet families. Only families with compact support (i.e. they are nonzero only in a finite interval) will be presented. All these wavelet families can be classified as either being an orthogonal or biorthogonal family. The meaning of orthogonality will be explained later. Each orthogonal wavelet family is characterized by two functions, the mother scaling function ϕ and the mother wavelet ψ. In the case of biorthogonal families one has a dual scaling function $\tilde{\phi}$ and a dual wavelet $\tilde{\psi}$ in addition to the non-dual quantities.

Figure 1 shows the orthogonal Haar wavelet family, which is conceptually the simplest wavelet family. It is too crude to be useful for any numerical work, but its simplicity will help us to illustrate some basic wavelet concepts. The Haar wavelet is identical to the zero-th degree Daubechies[3] wavelet.

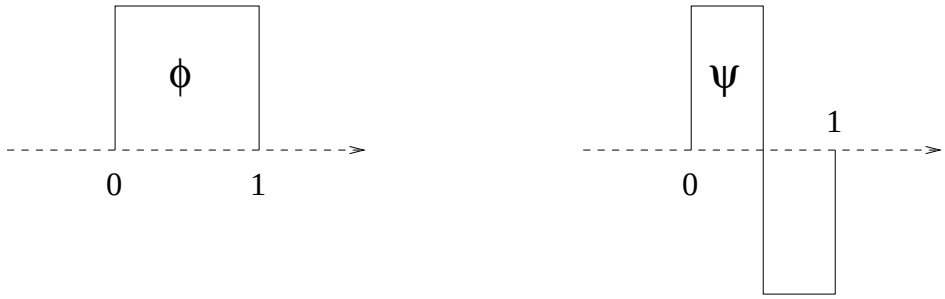


Figure 1. The Haar scaling function ϕ and wavelet ψ.

Figure 2 shows the 4 and 8 order Daubechies wavelets. Note that both the regularity and the support length increase with increasing order of the wavelets. The Daubechies family is an orthogonal wavelet family.

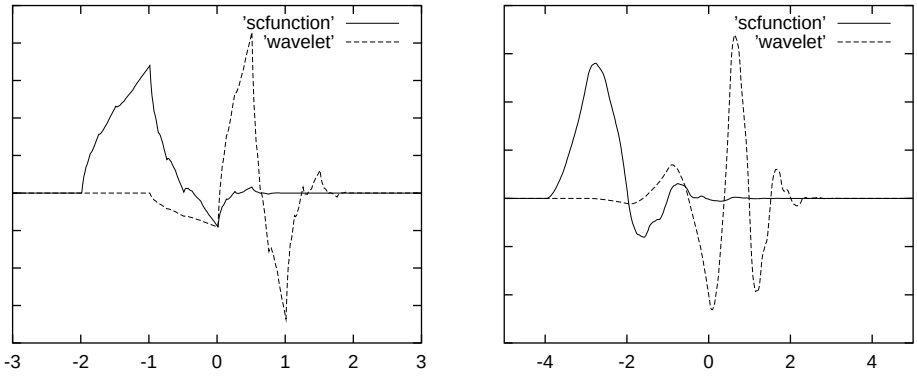


Figure 2. The orthogonal Daubechies scaling function and wavelet of degree 4 (left panel) and 8 (right panel).

Figure 3 shows a biorthogonal interpolating wavelet family of degree 4. It is smoother than other families of the same degree. Note that the scaling function vanishes at all integer points except at the origin.

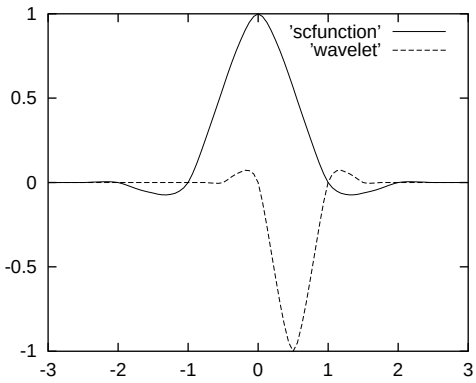


Figure 3. The interpolating scaling function and wavelet of degree 4.

3 Forming a Basis Set

To obtain a basis set at a certain resolution level k one can use all the integer translations of the mother scaling function of some wavelet family,

$$\phi_i^k(x) \propto \phi(2^k x - i)\,. \tag{1}$$

Note that with this convention higher resolution corresponds to larger values of k. Since high resolution scaling functions are skinnier, more translation indices i are allowed for a interval of fixed length. Some examples for an unspecified wavelet family are shown in Figure 4.

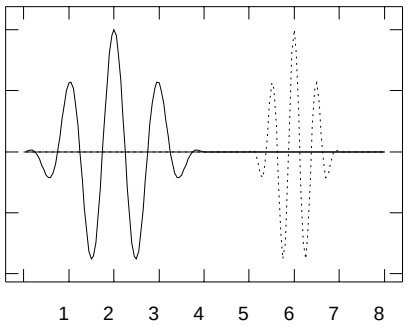


Figure 4. Two basis functions $\phi_2^0(x)$ (solid line) and $\phi_{12}^1(x)$ (dotted line) for an arbitrary wavelet family.

Exactly the same scaling and shifting operations can of course also be applied to the wavelets,

$$\psi_i^k(x) \propto \psi(2^k x - i)\,. \tag{2}$$

This set of wavelet basis functions can be added as a basis to the scaling functions as will be explained in the following.

4 The Haar Wavelet

In the case of the Haar family, any function which can exactly be represented at any level of resolution is necessarily piecewise constant. One such function is shown in Figure 5.

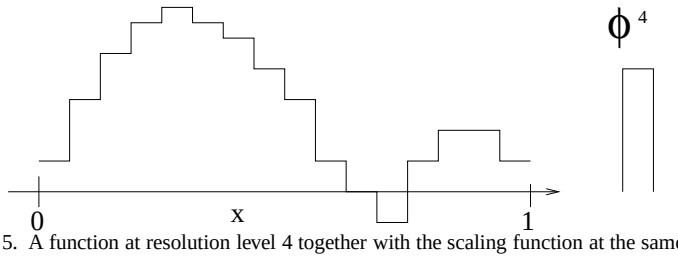


Figure 5. A function at resolution level 4 together with the scaling function at the same resolution level.

Evidently this function can be written as a linear combination of the scaling functions $\phi_i^4(x)$

$$f = \sum_{i=0}^{15} s_i^4\, \phi_i^4(x)\,, \tag{3}$$

where $s_i^4 = f(i/16)$.

Another, more interesting, possibility consists of expanding a function with respect to both scaling functions and wavelets of different resolution. Even though such a expansion contains both scaling functions and wavelets, we will refer to it as a wavelet representation to distinguish it from the our scaling function representation of Equation (3). A wavelet representation is possible because a scaling function at resolution level k is always a linear combination of a scaling function and a wavelet at the next coarser level $k-1$ as shown in Figure 6.

Using this relation depicted in Figure 6, we can write any linear combination of the two scaling functions $\phi_{2i}^k(x)$ and $\phi_{2i+1}^k(x)$ as a linear combination of $\phi_i^{k-1}(x)$ and $\psi_i^{k-1}(x)$. Hence we can write f as

$$f = \sum_{i=0}^{7} s_i^3\, \phi_i^3(x) + \sum_{i=0}^{7} d_i^3\, \psi_i^3(x)\,. \tag{4}$$

It is easy to verify that the transformation rule for the coefficients is given by

$$s_i^{k-1} = \frac{1}{2} s_{2i}^k + \frac{1}{2} s_{2i+1}^k \qquad ; \qquad d_i^{k-1} = \frac{1}{2} s_{2i}^k - \frac{1}{2} s_{2i+1}^k\,. \tag{5}$$

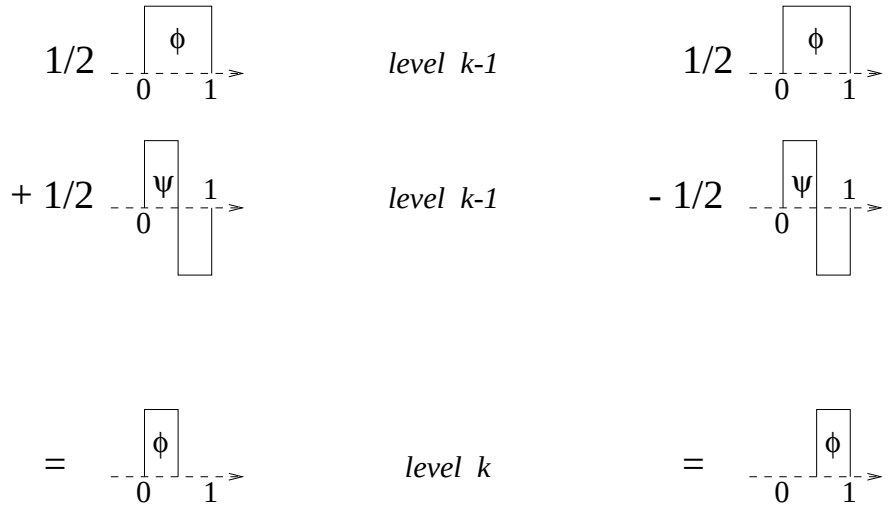


Figure 6. A skinny (level k) scaling function is a linear combination of a fat (level $k-1$) scaling function and a fat wavelet.

So to calculate the expansion coefficients with respect to the scaling functions at the next coarser level, we have to take an average over expansion coefficients at the finer resolution level. Because we have to take some weighted sum these coefficients are denoted by s. To get the expansion coefficients with respect to the wavelet, we have to take some weighted difference and the coefficients are accordingly denoted by d. The wavelet part contains mainly high frequency components and by doing this transformation we therefore peel off the highly oscillatory parts of the function. The remaining part represented by the coefficients s_i^{k-1} is therefore smoother. It is admittedly difficult to talk about smoothness for this kind of piecewise constant functions. This effect will be more visible for better wavelet families discussed later. For the case of our example in Figure 5 this remaining part after one transformation step is shown in Figure 7.

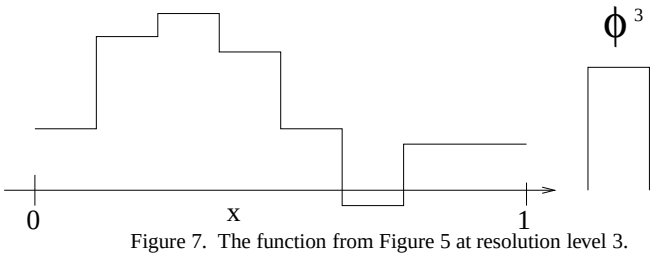


Figure 7. The function from Figure 5 at resolution level 3.

For any data set whose size is a power of 2, we can now apply this transformation repeatedly. In each step the number of s coefficients will be cut into half. So we have

to stop the procedure as soon as there is only one s coefficient left. Such a series of transformation steps is called a forward Haar wavelet transform. The resulting wavelet representation of the function in Equation (3) is then

$$f = s_0^0\phi_0^0(x) + d_0^0\psi_0^0(x) + \sum_{i=0}^{1} d_i^1\psi_i^1(x) + \sum_{i=0}^{3} d_i^2\psi_i^2(x) + \sum_{i=0}^{7} d_i^3\psi_i^3(x)\,. \qquad (6)$$

Note that in both cases we need exactly 16 coefficients to represent the function. In the coming sections such wavelet representations will be the focus of our interest.

By doing a backward wavelet transform, we can go back to the original scaling function representation of Equation (3). Starting at the coarsest resolution level, we have to express each scaling function and wavelet on the coarse level in terms of scaling functions at the finer level. This can be done exactly because wavelet families satisfy the so-called refinement relations depicted in Figure 8 for the Haar family.

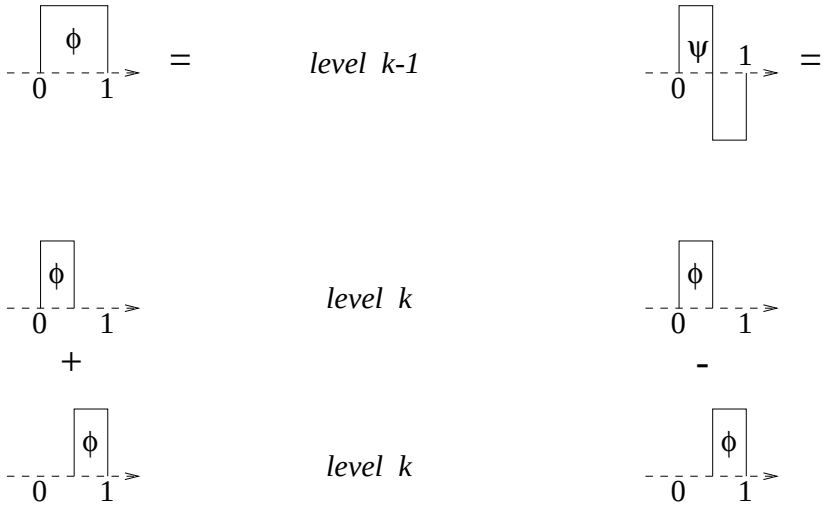


Figure 8. Fat (level $k-1$) scaling functions and fat wavelets are linear combinations of skinny (level k) scaling functions.

It then follows that we have to back-transform the coefficients in the following way

$$s_{2i}^{k+1} = s_i^k + d_i^k \qquad ; \qquad s_{2i+1}^{k+1} = s_i^k - d_i^k\,. \qquad (7)$$

5 The Concept of Multi-Resolution Analysis

In the previous sections a very intuitive introduction to wavelet theory was given. The formal theory behind wavelets is called Multi-Resolution Analysis[3] (MRA). Even though the formal definitions of MRA are usually not required for practical work, we will for completeness briefly present them. The equations which are useful for numerical work will be listed afterwards.

5.1 Formal definition of Multi-Resolution Analysis

- A Multi-Resolution Analysis consists of a sequence of successive approximation spaces V_k and associated dual spaces $\tilde{V}_k$, (which turn out to be the scaling function spaces and their dual counterpart) satisfying

$$V_k \subset V_{k+1} \qquad ; \qquad \tilde{V}_k \subset \tilde{V}_{k+1} \; .$$

- If a function $f(x)$ is contained in the space V_k, the compressed function $f(2x)$ has to be contained in the higher resolution space V_{k+1},

$$f(x) \in V_k \Leftrightarrow f(2x) \in V_{k+1} \qquad ; \qquad f(x) \in \tilde{V}_k \Leftrightarrow f(2x) \in \tilde{V}_{k+1} \; .$$

- If a function $f(x)$ is contained in the space V_k, its integer translate has to be contained in the same space,

$$f(x) \in V_0 \Leftrightarrow f(x+1) \in V_0 \qquad ; \qquad f(x) \in \tilde{V}_0 \Leftrightarrow f(x+1) \in \tilde{V}_0 \; .$$

- The union of all these spaces is the $L^2(\Re)$ space,

$$\overline{\bigcup_k V_k} = L^2(\Re) \; .$$

- There exists a biorthogonal pair of functions spanning V_k,

$$\int \tilde{\phi}_i^k(x) \, \phi_j^k(x) \, dx = \delta_{i,j} \; .$$

The wavelet spaces W_k, $\tilde{W}_k$ are then defined as the complement (orthogonal complement in the case of orthogonal families) of V_k in V_{k+1},

$$V_{k+1} = V_k \oplus W_k \qquad ; \qquad \tilde{V}_{k+1} = \tilde{V}_k \oplus W_k \; .$$

5.2 Basic formulas for biorthogonal wavelet families

The formal MRA requirements listed above lead to the following useful basic facts of wavelet analysis. The interested reader can find the nontrivial proofs of these formulas in the book by Daubechies[3].

- A biorthogonal wavelet family of degree m is characterized by 4 finite filters denoted by h_j, $\tilde{h}_j$, g_j, $\tilde{g}_j$. Since we will mainly deal with symmetric wavelet families, whose filters have a natural symmetry center, we will adopt a convention where the nonzero filter elements are in the interval $j = -m, ..., m$, and where m is even. In case the number of nonzero filter elements does not fit into this convention, it is always possible to pad the filters on both sides with zeroes, and to increase m artificially until it is compatible with this convention.

The filter coefficients satisfy the orthogonality relations

$$\sum_l h_{l-2i}\,\tilde{h}_{l-2j} = \delta_{i,j}\,, \tag{8}$$

$$\sum_l g_{l-2i}\,\tilde{g}_{l-2j} = \delta_{i,j}\,, \tag{9}$$

$$\sum_l h_{l-2i}\,\tilde{g}_{l-2j} = 0\,, \tag{10}$$

$$\sum_l \tilde{h}_{l-2i}\,g_{l-2j} = 0 \tag{11}$$

and the symmetry relations

$$g_{i+1} = (-1)^{i+1}\tilde{h}_{-i}\,, \tag{12}$$

$$\tilde{g}_{i+1} = (-1)^{i+1}h_{-i}\,. \tag{13}$$

- Scaling functions and wavelets at a coarse level can be written as the following linear combinations of scaling functions at a higher resolution level. These equations are called refinement relations,

$$\phi(x) = \sum_{j=-m}^{m} h_j\,\phi(2x-j)\,, \tag{14}$$

$$\psi(x) = \sum_{j=-m}^{m} g_j\,\phi(2x-j)\,, \tag{15}$$

$$\tilde{\phi}(x) = 2\sum_{j=-m}^{m} \tilde{h}_j\,\tilde{\phi}(2x-j)\,, \tag{16}$$

$$\tilde{\psi}(x) = 2\sum_{j=-m}^{m} \tilde{g}_j\,\tilde{\phi}(2x-j)\,. \tag{17}$$

In terms of the the two index multi level basis functions defined by,

$$\phi_i^k(x) = \phi(2^k x - i)\,, \tag{18}$$

$$\psi_i^k(x) = \psi(2^k x - i)\,, \tag{19}$$

$$\tilde{\phi}_i^k(x) = 2^k\tilde{\phi}(2^k x - i)\,, \tag{20}$$

$$\tilde{\psi}_i^k(x) = 2^k\tilde{\psi}(2^k x - i)\,, \tag{21}$$

the refinement relations are,

$$\phi_i^k(x) = \sum_{j=-m}^{m} h_j \, \phi_{2i+j}^{k+1}(x) \, , \tag{22}$$

$$\psi_i^k(x) = \sum_{j=-m}^{m} g_j \, \phi_{2i+j}^{k+1}(x) \, , \tag{23}$$

$$\tilde{\phi}_i^k(x) = \sum_{j=-m}^{m} \tilde{h}_j \, \tilde{\phi}_{2i+j}^{k+1}(x) \, , \tag{24}$$

$$\tilde{\psi}_i^k(x) = \sum_{j=-m}^{m} \tilde{g}_j \, \tilde{\phi}_{2i+j}^{k+1}(x) \, . \tag{25}$$

- A wavelet analysis (forward) transform is given by

$$s_i^{k-1} = \sum_{j=-m}^{m} \tilde{h}_j s_{j+2i}^k \, , \tag{26}$$

$$d_i^{k-1} = \sum_{j=-m}^{m} \tilde{g}_j s_{j+2i}^k \, .$$

A wavelet synthesis (backward) transform is given by

$$s_{2i}^{k+1} = \sum_{j=-m/2}^{m/2} h_{2j} \, s_{i-j}^k + g_{2j} \, d_{i-j}^k \tag{27}$$

$$s_{2i+1}^{k+1} = \sum_{j=-m/2}^{m/2} h_{2j+1} \, s_{i-j}^k + g_{2j+1} \, d_{i-j}^k \, .$$

These two equations are generalizations of equations (5), (7) that we derived in an intuitive way.

The wavelet transform is in principle for periodic data sets. Therefore the subscripts of the s and d coefficients have to be wrapped around once they are out of bounds.

- The fundamental functions satisfy the following orthogonality relations,

$$\int \tilde{\phi}_i^k(x) \phi_j^k(x) dx = \delta_{i,j} \, , \tag{28}$$

$$\int \tilde{\psi}_i^k(x) \phi_j^q(x) dx = 0 \, , \; k \geq q \, , \tag{29}$$

$$\int \psi_i^k(x) \tilde{\phi}_j^q(x) dx = 0 \, , \; k \geq q \, , \tag{30}$$

$$\int \psi_i^k(x) \tilde{\psi}_j^q(x) dx = \delta_{k,q} \delta_{i,j} \, . \tag{31}$$

5.3 Basic formulas for orthogonal wavelet families

- An orthogonal wavelet family of degree m is characterized by 2 finite filters denoted by h_j, g_j, satisfying the orthogonality relations

$$\sum_l h_{l-2i}\, h_{l-2j} = \delta_{i,j}\ , \tag{32}$$

$$\sum_l g_{l-2i}\, g_{l-2j} = \delta_{i,j}\ , \tag{33}$$

$$\sum_l h_{l-2i}\, g_{l-2j} = 0 \tag{34}$$

  and the symmetry relation

$$g_{i+1} = (-1)^{i+1} h_{-i}\ . \tag{35}$$

- The refinement relations are

$$\phi(x) = \sqrt{2} \sum_{j=-m}^{m} h_j\, \phi(2x-j)\ , \tag{36}$$

$$\psi(x) = \sqrt{2} \sum_{j=-m}^{m} g_j\, \phi(2x-j)\ . \tag{37}$$

  In terms of the the two index multi level basis functions defined by

$$\phi_i^k(x) = \sqrt{2}^k \phi(2^k x - i)\ , \tag{38}$$

$$\psi_i^k(x) = \sqrt{2}^k \psi(2^k x - i)\ , \tag{39}$$

  the refinement relations are

$$\phi_i^k(x) = \sum_{j=-m}^{m} h_j\, \phi_{2i+j}^{k+1}(x)\ , \tag{40}$$

$$\psi_i^k(x) = \sum_{j=-m}^{m} g_j\, \phi_{2i+j}^{k+1}(x)\ . \tag{41}$$

- The formulas for the forward and backward wavelet transforms are identical to the biorthogonal case (Equation (26) and (27)), with the exception that the filters $\tilde{h}$ and $\tilde{g}$ have to be replaced by the filters h and g in the forward transform.

- The fundamental functions satisfy the orthogonality relations,

$$\int \phi_i^k(x)\phi_j^k(x)dx = \delta_{i,j}\ , \tag{42}$$

$$\int \psi_i^k(x)\phi_j^q(x)dx = 0\ ,\ k \geq q\ , \tag{43}$$

$$\int \psi_i^k(x)\psi_j^q(x)dx = \delta_{k,q}\delta_{i,j}\ . \tag{44}$$

6 The Fast Wavelet Transform

One single sweep in a wavelet transformation (Eq. 26, Eq. 27) is a convolution with a short filter that can be done with linear scaling with respect to the size of the data set. An entire wavelet analysis transformation consists of several sweeps where in each consecutive sweep the amount of data to be transformed is cut into half. The total number of arithmetic operations is therefore given by a geometric series and is proportional to the data set. More precisely, if our filters h and g have length $2m$ the operation count is given by $2m(n+n/2+n/4+...) < 4mn$. The entire wavelet analysis scales therefore linearly. An entire wavelet synthesis is just the reverse operation and scales linearly as well. Below the evolution of a data set in a wavelet analysis is shown.

original data

$$s_0^4\ s_1^4\ s_2^4\ s_3^4\ s_4^4\ s_5^4\ s_6^4\ s_7^4\ s_8^4\ s_9^4\ s_{10}^4\ s_{11}^4\ s_{12}^4\ s_{13}^4\ s_{14}^4\ s_{15}^4\ =\ S^4$$

after first sweep

$$s_0^3\ s_1^3\ s_2^3\ s_3^3\ s_4^3\ s_5^3\ s_6^3\ s_7^3\ d_0^3\ d_1^3\ d_2^3\ d_3^3\ d_4^3\ d_5^3\ d_6^3\ d_7^3\ =\ S^3, D^3$$

after second sweep

$$s_0^2\ s_1^2\ s_2^2\ s_3^2\ d_0^2\ d_1^2\ d_2^2\ d_3^2\ d_0^3\ d_1^3\ d_2^3\ d_3^3\ d_4^3\ d_5^3\ d_6^3\ d_7^3\ =\ S^2, D^2, D^3$$

after third sweep

$$s_0^1\ s_1^1\ d_0^1\ d_1^1\ d_0^2\ d_1^2\ d_2^2\ d_3^2\ d_0^3\ d_1^3\ d_2^3\ d_3^3\ d_4^3\ d_5^3\ d_6^3\ d_7^3\ =\ S^1, D^1, D^2, D^3$$

final data

$$s_0^0\ d_0^0\ d_0^1\ d_1^1\ d_0^2\ d_1^2\ d_2^2\ d_3^2\ d_0^3\ d_1^3\ d_2^3\ d_3^3\ d_4^3\ d_5^3\ d_6^3\ d_7^3\ =\ S^0, D^0, D^1, D^2, D^3$$

Note that this transformation from the original data to the final data corresponds exactly to the transformation done in an intuitive way to get from Equation (3) to Equation (6).

7 Interpolating Wavelets

As will be discussed later, interpolating wavelets have many properties, which make them highly suitable as basis sets for partial differential equations. At the same time they are conceptionally the simplest wavelets. We will therefore describe the construction of the elementary interpolating wavelet[7,6] in detail.

The construction of interpolating wavelets is closely connected to the question of how to construct a continuous function $f(x)$ if only its values f_i on a finite number of grid points i are known. One way to do this is by recursive interpolation. In a first step we interpolate the functional values on all the midpoints by using for instance the functional values of two grid points to the right and of two grid points to the left of the midpoint. These four functional values allow us to construct a third order polynomial and we can then evaluate it at the midpoint. In the next step, we take this new data set, which is now twice as large as the original one, as the input for a new midpoint interpolation procedure. This can be done recursively ad infinitum until we have a quasi continuous function.

Let us now show, how this interpolation prescription leads to a set of basis functions. Denoting by the Kronecker δ_{i-j} a data set whose elements are all zero with the exception of the element at position j, we can write any initial data set as a linear combination of such Kronecker data sets: $f_i = \sum_j f_j \delta_{i-j}$. Now the whole interpolation procedure is clearly linear, i.e. the sum of two interpolated values of two separate data sets is equal to the interpolated value of the sum of these two data sets. This means that we can instead also take all the Kronecker data sets as the input for separate ad-infinitum interpolation procedures, to obtain a set of functions $\phi(x - j)$. The final interpolated function is then identical to $f(x) = \sum_j f_j \phi(x - j)$. If the initial grid values f_i were the functional values of a polynomial of degree less than four, we obviously will have exactly reconstructed the original function from its values on the grid points. Since any smooth function can locally be well approximated by a polynomial, these functions $\phi(x)$ are good basis functions and we will use them as scaling functions to construct a wavelet family.

The first construction steps of an interpolating scaling function are shown in Figure 9 for the case of linear interpolation. The initial Kronecker data set is denoted by the big dots. The additional data points obtained after the first interpolation step are denoted by medium size dots and the additional data points obtained after the second step by small dots.

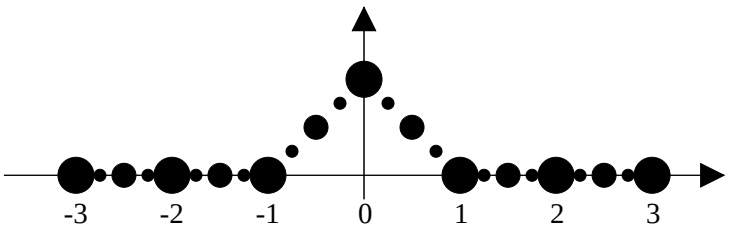


Figure 9. The first two steps of a recursive interpolation procedure in the case of simple linear interpolation. The original data point are represented by the big dots, data points filled in by the following two interpolation steps by medium and small dots.

Continuing this process ad infinitum will then result in the function shown in the left panel of Figure 10. If an higher order interpolation scheme is used the function shown in the right panel of Figure 10 is obtained.
By construction it is clear, that $\phi(x)$ has compact support. If an $(m-1)$-th order interpolation scheme is used, the filter length is $(m-1)$ and the support interval of the scaling function is $[-(m-1); (m-1)]$.

It is also not difficult to see that the function $\phi(x)$ satisfies the refinement relation. Let us again consider the interpolation ad infinitum of a Kronecker data set which has everywhere zero entries except at the origin. We can now split up this process into the first step where we calculate the half-integer grid point values and a remaining series of separate ad infinitum interpolations for all half-integer Kronecker data sets, which are necessary to represent the data set obtained by the first step. Doing the ad-infinitum interpolation for a half integer data set with a unit entry at (half integer) position j, we obviously obtain the same scaling function, just compressed by a factor of 2, $\phi(2x - j)$. If we are using a $(m-1)$-th order interpolation scheme (i.e. m input data for the interpolation process) we

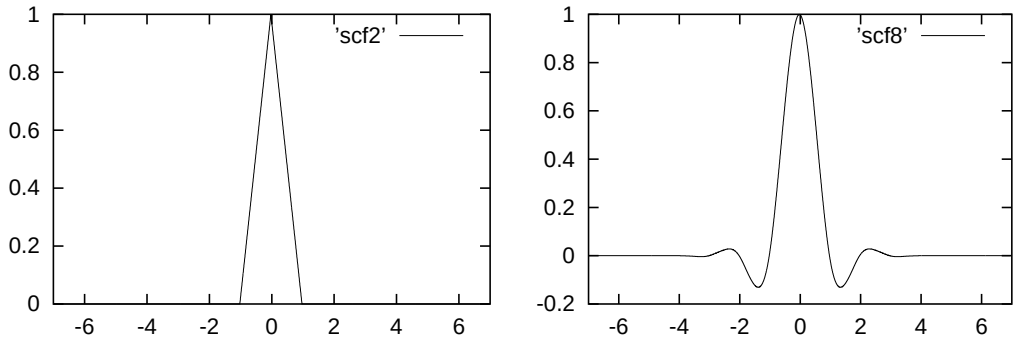


Figure 10. A Kronecker delta interpolated ad infinitum with linear interpolation (left panel) and 7-th order interpolation (right panel) .

thus get the relation

$$\phi(x) = \sum_{j=-m+1}^{m-1} \phi(j/2)\ \phi(2x-j)\ . \tag{45}$$

Comparing this equation with the refinement relation Equation (14) we can identify the first filter h as

$$h_j = \phi(j/2)\ ,\ \ j = -m+1, m-1\ .$$

For the case of third order interpolation the numerical values of h follow from the standard interpolation formula and are given by:

```
h =   {-1/16 , 0 , 9/16 , 1 , 9/16 , 0 , -1/16}
j =      -3    -2    -1    0     1    2     3
```

Let us next determine the filter $\tilde{h}$. Let us consider a function $f(x)$ which is band-limited in the wavelet sense, i.e which can exactly be represented by a superposition of scaling functions at a certain resolution level K

$$f(x) = \sum_j c_j \phi_j^K(x)\ .$$

It then follows from the orthogonality relation Equation (28) that

$$c_j = \int \tilde{\phi}_j^K(x) f(x) dx\ .$$

Now we have seen above that with respect to interpolating scaling functions, a band-limited function is just any polynomial of degree less than or equal to $m-1$, and that in this case the expansion coefficients c_j are just the functional values at the grid points (Equation (45)). We therefore have

$$\int \tilde{\phi}_j^K(x) f(x) dx = f_j\ ,$$

which shows that the dual scaling function $\tilde{\phi}$ is the delta function. Obviously the delta function satisfies a trivial refinement relation $\delta(x) = 2\delta(2x)$ and from Equation (16) we conclude that $\tilde{h}_j = \delta_j$.

```
ht =   {  0 ,  0  , 1 ,  0  , 0  }
j  =      -2   -1    0    1    2
```

From the symmetry Equations (12), (13) for the filters we can now determine the two remaining filters and we have thus completely specified our wavelet family. For $\tilde{g}_j$ we obtain

```
gt =   {  0 , 0 , -1/16 , 0 , 9/16 , -1 , 9/16 , 0 , -1/16 }
j  =      -4  -3    -2    -1    0      1    2     3     4
```

For g_j we obtain

```
g =   {  0 ,  0 ,  0 , -1 ,  0 }
j =      -2   -1    0    1    2
```

As expected, these 4 filters satisfy the orthogonality conditions (8) to (11).

Due to the easy structure of the filters in this case, the backward transform can be done by inspection to obtain the form of the 4 fundamental functions ϕ, ψ, $\tilde{\phi}$ and $\tilde{\psi}$. In the case of the scaling function ϕ, the d coefficients at all resolution levels vanish. For the even elements Equation (27) becomes

$$s_{2i}^{k+1} = s_i^k \; .$$

So the even grid points on the finer grid take on the values of the coarser grid. The odd filter elements are just the interpolating coefficients giving:

$$s_{2i+1}^{k+1} = h_3 \, s_{i-1}^k + h_1 \, s_{i+0}^k + h_{-1} \, s_{i+1}^k + h_{-3} \, s_{i+2}^k \; .$$

So the values at the odd fine grid points are just interpolated from the coarse grid points. In summary we thus see that an infinite series of backward transforms just describes the ad-infinitum interpolation process depicted in Figure 9.

In the case of the wavelet ψ the only nonzero d coefficient in the input data will generate in the first step a s data set where again only one coefficient is nonzero, since the g filter has only one nonzero entry. Continuing the procedure one will thus obtain for the wavelet a negative scaling function compressed by a factor of 2, $\psi(x) = -\phi(2x-1)$.

To generate the dual functions $\tilde{\phi}$ and $\tilde{\psi}$, one has to replace the filters h and g in the backward transform by the dual counterparts $\tilde{h}$ and $\tilde{g}$. For the case of the dual scaling function $\tilde{\phi}$, one sees by inspection that the backward transform equations Equation (27) become:

$$s_{2i+1}^{k-1} = 0 \qquad ; \qquad s_{2i}^{k-1} = \begin{cases} 1 \text{ if } i = 0 \\ 0 \text{ otherwise} \end{cases}$$

As one should, one thus obtains a delta function

$$\tilde{\phi}(x) = \delta(x) \; . \tag{46}$$

For the case of a dual wavelet $\tilde{\psi}$ the argument is analogous to the non-dual case. In the first step of the backward transform the filter $\tilde{g}$ generates 5 nonzero s coefficients, which will become 5 delta functions through the action of the filter $\tilde{h}$. We get

$$\begin{aligned} \tilde{\psi}(x) = & -\tfrac{1}{16}\delta((x-\tfrac{1}{2})+3/2) + \tfrac{9}{16}\delta((x-\tfrac{1}{2})+1/2) - \delta((x-\tfrac{1}{2})) \\ & + \tfrac{9}{16}\delta((x-\tfrac{1}{2})+1/2) - \tfrac{1}{16}\delta((x-\tfrac{1}{2})+3/2) \; . \end{aligned} \tag{47}$$

We thus see that the interpolating wavelet is a very special case in that its scaling function and wavelet have the same functional form and that the dual functions are related to the delta function. The non-dual functions are shown in Figure 3 . Filters for interpolating wavelets of other degrees are given in the Appendix.

8 Expanding Polynomials in a Wavelet Basis

Functions of practical interest are of course not simple polynomials, and it will be discussed later how to expand arbitrary functions in a wavelet basis. For several proofs the expansion of polynomials in a wavelet basis is however important and we will therefore derive the following theorem: The scaling function expansion coefficients $s_i(l)$ of a polynomial of degree l are themselves a polynomial of the same degree l.

Let us first demonstrate the theorem for the trivial case of a constant, i.e. $l = 0$. The expansion coefficients $s_i(0)$ are given by $\int \tilde{\phi}(x-i)dx$. Assuming $\int \tilde{\phi}(x)dx$ is normalized to 1 we thus obtain $s_i(0) = 1$.

In the linear case (i.e. $l = 1$) we have $s_i(1) = \int \tilde{\phi}(x-i)xdx$. For the shifted coefficient we get

$$\begin{aligned} s_{i+1}(1) &= \int \tilde{\phi}(x-i-1)xdx = \int \tilde{\phi}(x-i)(x+1)dx \\ &= s_i(1) + 1 \; . \end{aligned} \tag{48}$$

So we see that $s_i(1)$ satisfies the difference equation for a linear polynomial and that it is therefore a linear polynomial.

For arbitrary degree l we get

$$\begin{aligned} s_{i+1}(l) &= \int \tilde{\phi}(x-i-1)x^l dx = \int \tilde{\phi}(x-i)(x+1)^l dx \\ &= \sum_\tau \int \tilde{\phi}(x-i)\frac{l!}{\tau!(l-\tau)!}x^\tau dx \\ &= \sum_\tau \frac{l!}{\tau!(l-\tau)!}s_i(\tau) \; . \end{aligned} \tag{49}$$

So wee see indeed that $s_i(l)$ is a polynomial of lth degree since it satisfies the corresponding difference equation, which proves the theorem.

9 Orthogonal Versus Biorthogonal Wavelets

The interpolating wavelets constructed above are a special case of so-called biorthogonal wavelet families. The interpolating wavelets have the property of being the smoothest ones for a fixed filter length. On the other hand the dual functions of the interpolating wavelet family are the least smooth ones. Loosely speaking the sum of the smoothness of the dual and non-dual space are a constant for a given filter length. For a given filter length one can therefore either go for maximum smoothness in the dual or non-dual space. The interpolating wavelets are your favorite choice if you want maximum smoothness in the non-dual space.

Wavelets are called orthogonal if the dual quantities are equal to the non-dual quantities. In the case of orthogonal wavelets the smoothness in dual and non-dual space is thus obviously the same. They are therefore not as smooth as the interpolating wavelets. The smoothness properties of the Daubechies family are actually not as bad as on might expect from looking at the "ugly" plots. With the 4-th order family one can exactly represent linear function, with the 6-th order family quadratic and with the 8-th order family cubic polynomials.

10 Expanding Functions in a Wavelet Basis

As we have seen, there are two possible representations of a function within the framework of wavelet theory. The first one is called scaling function representation and involves only scaling functions. The second is called wavelet representation and involves wavelets as well as scaling functions. Both representations are completely equivalent and exactly the same number of coefficients are needed. The scaling function representation is given by

$$f(x) = \sum_j s_j^{Kmax} \phi_j^{Kmax}(x) \; . \tag{50}$$

Evidently this approximation is more accurate if we use skinnier scaling functions from a higher resolution level $Kmax$. From the orthogonality relations (28) it follows, that the coefficients are given by

$$s_j^{Kmax} = \int \tilde{\phi}_j^{Kmax}(x) \; f(x) dx \; . \tag{51}$$

Once we have a set of coefficients s_j^{Kmax} we can use a full forward wavelet transform to obtain the wavelet representation

$$f(x) = \sum_j s_j^{Kmin} \; \phi_j^{Kmin}(x) + \sum_{K=Kmin}^{Kmax} \sum_j d_j^K \; \psi_j^K(x) \; . \tag{52}$$

Alternatively, one could also directly calculate the d coefficients by integration

$$d_j^K = \int \tilde{\psi}_j^K(x) \; f(x) dx \; . \tag{53}$$

The above Equation (53) follows from the orthogonality relations (29) to (31). So we see that if we want to expand a function either in scaling functions or wavelets, we have to perform integrations at some point to calculate the coefficients. For general wavelet families this integration is fairly cumbersome[5] and requires especially in 2 and 3 dimensions a substantial number of integration points. Furthermore it is not obvious how to do the integration if the function is only given in tabulated form. If one wants to obtain the same number of coefficients as one has functional values, one could either first interpolate the function to obtain the necessary number of integration points, which will introduce additional approximations. If one does not generate additional integration points, then the number of coefficients will necessarily be less than the number of functional values and information is thus lost. The interpolating wavelets discussed above are the glorious exception. Since the dual scaling function is a delta function and since the dual wavelet is a sum of the delta functions, one or a few data points are sufficient to do the integration exactly. In the case of periodic data sets, the filters will wrap around for data points close enough to the boundary of the periodic volume. One will therefore get exactly the same number of coefficients as one has data points and one has an invertible one-to-one mapping between the functional values on the grid and the expansion coefficients.

Non-periodic data sets can also be handled. In this case we have to put the non-periodic data set into a larger periodic data set consisting of zeroes. This composite data set will then contain the nonzero non-periodic data set in the middle surrounded by a layer of zeroes on all sides. If this layer of zeroes is broader than half the filter length $m/2$ opposite

ends will not interfere during one sweep of a wavelet transform and one obtains the correct representation of the non-periodic function. Correct means in this context that the value of the nonzero coefficients would not change if we made the surrounding layer of zeroes broader.

The interpolating wavelets are also unbeatable from the point of view of accuracy. The accuracy of a scaling function expansion depends on the smoothness of the scaling function. This is easy to see for the case of interpolating wavelets. The functional value of a scaling function expansion at any midpoint is given by interpolation and thus the error is also given by the well known interpolation error formula. If h is the grid spacing and $(m-1)$-th order interpolation is used then the error is proportional to h^m. So since the interpolating wavelets are the smoothest wavelets they are also the most accurate ones. If on the other hand one is willing to accept a certain error then the interpolating wavelets will meet this error criteria with the smallest number of expansion coefficients.

This fact can also be understood from a different point of view. Let us introduce the moments $\tilde{M}_l$,

$$\tilde{M}_l = \int \tilde{\psi}(x)\, x^l dx \; .$$

Now we know, that locally any smooth function can be approximated by a polynomial. Let us for simplicity consider the coefficient d_0^K at the origin,

$$d_0^K = \sum_{\nu=0}^{\infty} \int f^\nu(0) \frac{x^\nu}{\nu!}\, \tilde{\psi}_0^K(x) dx \; .$$

If the first L moments $l = 0, ..., L-1$ vanish this becomes

$$d_0^K = \sum_{\nu=L}^{\infty} h^\nu C_\nu \; ,$$

where we have used the fact that $\tilde{\psi}$ is a sum of delta functions and where C_ν are appropriate constants. The d coefficients decay therefore as h^L and since the error is proportional to the coefficients of the wavelets which are discarded, the error is proportional to h^L as well. In the case of the 4-th order interpolating wavelet it is easy to see, that the first 4 moments vanish, $\tilde{M}_l = 0, l = 0, 1, 2, 3$ and thus the error is indeed proportional to h^4. The measured decay of the d coefficients for the case of a Gaussian is shown in Figure 11.

This relation between the error in the expansion of a function and the number of vanishing moments is not only valid for interpolating wavelets but also holds true for other wavelet families.

11 Wavelets in 2 and 3 Dimensions

The easiest way to construct a wavelet basis in higher dimensional spaces is by forming product functions[3]. For simplicity of notation we will concentrate on the 2-dimensional case, the generalization to higher dimensional spaces being obvious. The space of all scaling functions of resolution level k is given in the 2-dimensional case by

$$\phi_{i1,i2}^k(x,y) = \phi_{i1}^k(x)\phi_{i2}^k(y) \; . \tag{54}$$

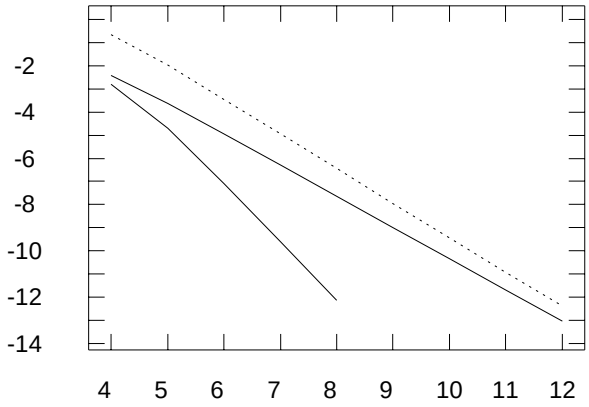


Figure 11. The decay of the d coefficients as a function of their resolution level on a double logarithmic scale. The solid lines show the result for a 4-th and 8-th order interpolating wavelet, the dashed line is for the 8-th order Daubechies family.

The wavelets consist of three types of products

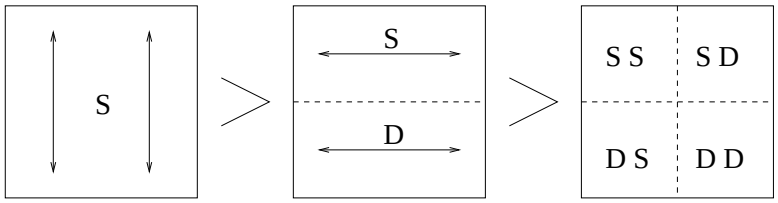
$$\psi[sd]^k_{i1,i2}(x,y) = \phi^k_{i1}(x)\psi^k_{i2}(y)\,, \tag{55}$$

$$\psi[ds]^k_{i1,i2}(x,y) = \psi^k_{i1}(x)\phi^k_{i2}(y)\,, \tag{56}$$

$$\psi[dd]^k_{i1,i2}(x,y) = \psi^k_{i1}(x)\psi^k_{i2}(y)\,. \tag{57}$$

In a 3-dimensional space the scaling functions are correspondingly of $[sss]$ type and there are 7 different classes of wavelets denoted by $[ssd]$, $[sds]$, $[sdd]$, $[dss]$, $[dsd]$, $[dds]$ and $[ddd]$.

It is easy to see, that both the many-dimensional scaling functions and wavelets satisfy refinement and orthogonality relations that are obvious generalizations of the 1-dimensional case. A wavelet transform step in the 2-dimensional setting is done by first transforming along the x and then along the y direction (or vice versa) as shown below.



To do a full 2-dim wavelet analysis one has to do a series of analysis steps. In each step the size of the active data set is reduced by 1/4 as shown in Figure 12. The total numerical effort therefore scales again linearly as in the one-dimensional case.

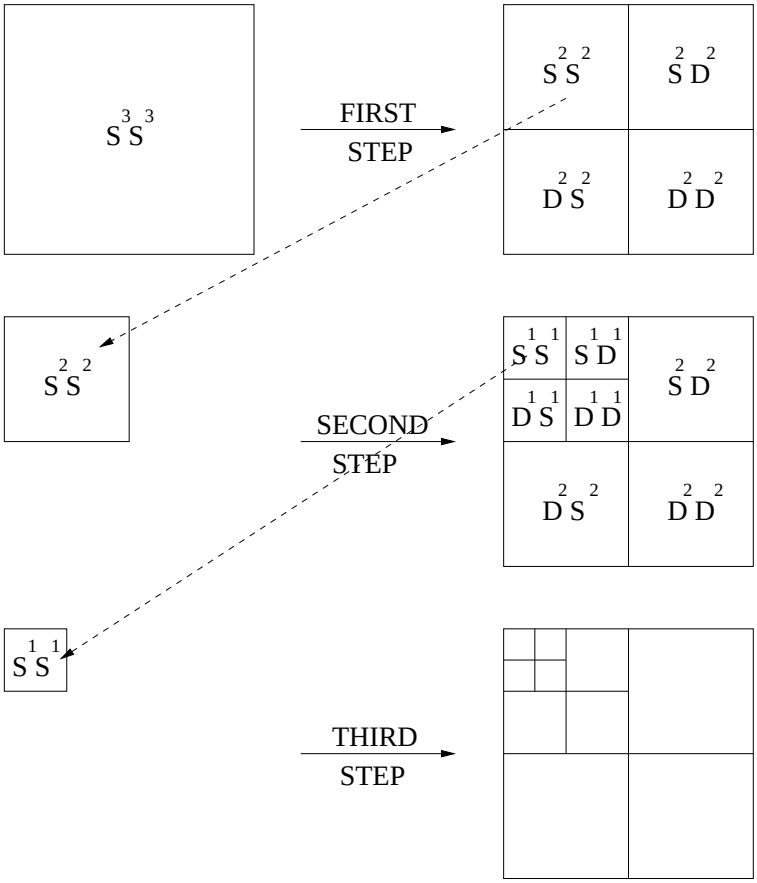


Figure 12. A full 2-dim wavelet analysis transformation.

12 Calculation of Differential Operators

As we have seen in the preceding chapter we need the matrix elements

$$\int \tilde{\phi}_i^k(x) \frac{\partial^l}{\partial x^l} \phi_j^k(x)\, dx\,, \tag{58}$$

$$\int \tilde{\psi}_i^k(x) \frac{\partial^l}{\partial x^l} \phi_j^k(x)\, dx\,, \tag{59}$$

$$\int \tilde{\phi}_i^k(x) \frac{\partial^l}{\partial x^l} \psi_j^k(x)\, dx\,, \tag{60}$$

$$\int \tilde{\psi}_i^k(x) \frac{\partial^l}{\partial x^l} \psi_j^k(x)\, dx\,, \tag{61}$$

to set up the SS, DS, SD and DD parts of the non-standard operator form. Matrix elements on different resolution levels k are obviously related by simple scaling relations. For instance

$$\int \tilde{\phi}_i^{k+1}(x)\frac{\partial^l}{\partial x^l}\phi_j^{k+1}(x)dx = 2^l \int \tilde{\phi}_i^k(x)\frac{\partial^l}{\partial x^l}\phi_j^k(x)dx \; . \tag{62}$$

So we just have to calculate these 4 matrix elements for one resolution level. On a certain resolution level, we can use the refinement relations to express the matrix elements involving wavelets in terms of matrix elements involving scaling functions (at a higher resolution level) only. Denoting the basic integral by a_i, where

$$a_i = \int \tilde{\phi}(x)\frac{\partial^l}{\partial x^l}\phi(x-i)dx \; , \tag{63}$$

we obtain

$$\int \tilde{\phi}_i(x)\frac{\partial^l}{\partial x^l}\phi_j(x)dx = a_{j-i} \; , \tag{64}$$

$$\int \tilde{\psi}_i(x)\frac{\partial^l}{\partial x^l}\phi_j(x)dx = 2^l \sum_{\nu,\mu} \tilde{g}_\nu h_\mu a_{2j-2i+\mu-\nu} \; , \tag{65}$$

$$\int \tilde{\phi}_i(x)\frac{\partial^l}{\partial x^l}\psi_j(x)dx = 2^l \sum_{\nu,\mu} \tilde{h}_\nu g_\mu a_{2j-2i+\mu-\nu} \; , \tag{66}$$

$$\int \tilde{\psi}_i(x)\frac{\partial^l}{\partial x^l}\psi_j(x)dx = 2^l \sum_{\nu,\mu} \tilde{g}_\nu g_\mu a_{2j-2i+\mu-\nu} \; . \tag{67}$$

To calculate a_i we follow Beylkin[9]. Using the refinement relations Equations (14) and (16) for ϕ and $\tilde{\phi}$ we obtain

$$\begin{aligned} a_i &= \int \tilde{\phi}(x)\frac{\partial^l}{\partial x^l}\phi(x-i)dx \qquad (68)\\ &= \sum_{\nu,\mu} 2\tilde{h}_\nu h_\mu \int \tilde{\phi}(2x-\nu)\frac{\partial^l}{\partial x^l}\phi(2x-2i-\mu)dx \\ &= \sum_{\nu,\mu} 2\tilde{h}_\nu h_\mu 2^{l-1} \int \tilde{\phi}(y-\nu)\frac{\partial^l}{\partial y^l}\phi(y-2i-\mu)dy \\ &= \sum_{\nu,\mu} \tilde{h}_\nu h_\mu 2^l \int \tilde{\phi}(y)\frac{\partial^l}{\partial y^l}\phi(y-2i-\mu+\nu)dy \\ &= \sum_{\nu,\mu} \tilde{h}_\nu h_\mu \; 2^l \; a_{2i-\nu+\mu} \end{aligned}$$

We thus have to find the eigenvector $\mathbf{a}$ associated with the eigenvalue of 2^{-l},

$$\sum_j A_{i,j} \; a_j = \left(\frac{1}{2}\right)^l a_i \; , \tag{69}$$

where the matrix $A_{i,j}$ is given by

$$A_{i,j} = \sum_{\nu,\mu} \tilde{h}_\nu h_\mu \, \delta_{j,2i-\nu+\mu} \; . \tag{70}$$

As it stands this eigensystem has a solution only if the rang of the matrix $A - 2^{-l} I$ is less than its dimension. For a well defined differential operator, i.e if l is less than the degree of smoothness of the scaling function this will be the case (for the 4-th order interpolating wavelet family the second derivative is for instance not defined). The system (69) is numerically unstable and it is therefore better to solve it using symbolic manipulations with a software such as Mathematica instead of solving it numerically.

The system of equations (69) determines the a_j's only up to a normalization factor. In the following, we will therefore derive the normalization equation. For simplicity, we will give the derivation only for the case of interpolating polynomials, even though the final result Equation (73) will hold in the general case.

From the normalization of the scaling function and from elementary calculus, it follows that

$$\int \phi(x) \frac{\partial^l}{\partial x^l} x^l \, dx = \int \phi(x) \; l! \, dx = l! \; . \tag{71}$$

On the other hand we know, that we can expand any polynomial of low enough degree exactly with the interpolating polynomials. The expansion coefficients are just i^l. So we obtain

$$\int \phi(x) \frac{\partial^l}{\partial x^l} \sum_i i^l \phi(x - i) = \sum_i i^l a_i \; . \tag{72}$$

By comparing Equation (71) and (72) we thus obtain the normalization condition

$$\sum_i i^l a_i = l! \; . \tag{73}$$

The interpolating wavelet family offers also an important advantage for the calculation of differential operators. Whereas in general derivative filters extend over the interval $[-2m; 2m]$ most of the border elements of interpolating wavelets are zero and their effective filter length is only $[-m + 2; m - 2]$.

Derivative filter coefficients for several families are listed in the Appendix.

13 Differential Operators in Higher Dimensions

As was pointed out before, higher dimensional wavelets can be constructed as products of one dimensional wavelets. The matrix elements of differential operators can therefore easily be derived.

Let us consider as an example the matrix elements of $\frac{\partial}{\partial x}$ with respect to the 2-dimensional scaling functions,

$$\int \tilde{\phi}^k_{i1,i2}(x,y) \frac{\partial}{\partial x} \phi^k_{j1,j2}(x,y) = \int \tilde{\phi}^k_{i1}(x) \tilde{\phi}^k_{i2}(y) \frac{\partial}{\partial x} \phi^k_{j1}(x) \phi^k_{j2}(y) = \delta_{i2-j2} a_{j1-i1} \; .$$

The remaining matrix elements among the wavelets and scaling functions can be derived along the same lines. Obviously a differential operator acting on x will only couple functions which have the same dependence with respect to y as indicated in Figure 13.

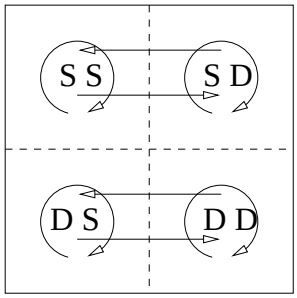


Figure 13. The coupling of the expansion coefficients under the action of a differential operator acting along the (horizontal) x axis.

14 The Solution of Poisson's Equation

In the following, a method[11] will be presented that uses interpolating scaling functions to solve Poisson's equation with free boundary conditions and $Nlog(N)$ scaling. The input is a charge density $\rho i1, i2, i3$ on a equally spaced 3-dimensional grid of $N = n_1 n_2 n_3$ grid points. For simplicity we put the grid spacing equal to 1. Since for interpolating scaling functions the expansion coefficients are just the values on the grid we can obtain from our discrete data set $\rho_{i1,i2,i3}$ a continuous charge distribution $\rho(\mathbf{r})$

$$\rho(\mathbf{r}) = \sum_{i1,i2,i3} \rho_{i1,i2,i3}\, \phi(x/h - i1)\, \phi(y/h - i2)\, \phi(z/h - i3) \tag{74}$$

It is not very difficult to prove that the discrete and continuous monopoles, i.e

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz\ \rho(\mathbf{r}) = \sum_{i1,i2,i3} \rho_{i1,i2,i3}$$

In the same way the discrete and continuous dipoles are identical, i.e

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz\ z\, \rho(\mathbf{r}) = \sum_{i1,i2,i3} i3\, \rho_{i1,i2,i3}$$

The potential on the grid point $j1, j2, j3$ (of same grid that was used for the input charge density) is then given by

$$\begin{aligned} V_{j1,j2,j3} &= \sum_{i1,i2,i3} \rho_{i1,i2,i3} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz \frac{\phi(x - i1)\, \phi(y - i2)\, \phi(z - i3)}{\sqrt{(x - j1)^2 + (y - j2)^2 + (z - j3)^2}} \\ &= \sum_{i1,i2,i3} \rho_{i1,i2,i3} F_{i1-j1,i2-j2,i3-j3} \end{aligned} \tag{75}$$

F is a long filter which depends only on the distance $i - j$ between the observation point j and the source point i. Since the above expression for the potential $V_{j1,j2,j3}$ is a convolution it can be calculated with FFT techniques at the cost of $N^3 \log(N^3)$ operations where N^3 is the number of grid points. It remains to calculate the values of the filter $F_{i1-j1,i2-j2,i3-j3}$. Calculating each of the N^3 filter elements as a 3-dimensional numerical integral would be too costly. The calculation becomes however feasible if the $1/r$

kernel is made separable. This can be achieved by representing it as a sum of Gaussians. The representation is best based on the identity

$$\frac{1}{r} = \frac{2}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-r^2 \exp(2s)+s} ds$$

Discretizing this integral we obtain

$$\frac{1}{r} = \sum_l w_l e^{-\gamma_l r^2} \tag{76}$$

With 89 well optimized values for w_l and γ_l it turns out that $1/r$ can be represented in the interval from 10^{-9} to 1 with an relative error of 10^{-8}. The 3-dimensional integral in Eq. 75 becomes then a sum of 89 terms each of which is a product of 1-dimensional integrals.

$$\int dx \int dy \int dz \frac{\phi(x-i1)\,\phi(y-i2)\,\phi(z-i3)}{\sqrt{(x-j1)^2+(y-j2)^2+(z-j3)^2}} =$$

$$\sum_{l=1}^{89} w_l \int dx \int dy \int dz\, \phi(x-i1)\,\phi(y-i2)\,\phi(z-i3)\, e^{-\gamma_l\left((x-j1)^2+(y-j2)^2+(z-j3)^2\right)} =$$

$$\sum_{l=1}^{89} w_l \int dx\, \phi(x-i1) e^{-\gamma_l (x-j1)^2} \int dy\, \phi(y-i2) e^{-\gamma_l (y-j2)^2} \int dz\, \phi(z-i3) e^{-\gamma_l (z-j3)^2}$$

Using 89 terms in Eq. 76 we have thus to solve just $89N$ one-dimensional integrals which can be done extremely rapidly on a modern computer. The main cost are thus the FFT's required to calculate the convolution with the kernel $F_{i1-j1,i2-j2,i3-j3}$.

The above presented method does not exploit the possibility to have adaptivity in a wavelet basis. Adpative methods to solve Poisson's equation on grids where the resolution varies by several orders of magnitude exist[10] as well. They are however based on more advanced concepts[4] such as non-standard operator forms and lifted interpolating wavelets.

15 The Solution of Schrödinger's Equation

Since the different Kohn-Sham orbitals in a density functional calculation have to be orthogonal, orthogonalization steps occur frequently in such a calculation. As a matter of fact these orthogonalization operations have cubic scaling and dominate thus the whole calculation for large system. It is therefore important that these operations can be done efficiently. This strongly suggests to use orthogonal Daubechies scaling functions and wavelets as basis functions for the Kohn-Sham orbitals. In spite of the striking advantages of Daubechies wavelets, the initial exploration of this basis set[8] did not lead to any algorithm that would be useful for real electronic structure calculations. This was due to the fact that an accurate evaluation of the local potential energy is difficult in a Daubechies wavelet basis. The kinetic energy part on the other hand is easy since it is just given by the Laplace operator. How to treat the Laplace operator has already been discussed. The obstacles in the evaluation of the potential energy have been overcome[12] recently and it was consequently shown that wavelets are an efficient basis set for electronic structure calculations[13] which outperforms plane waves for open structures. We will next discuss how

the different parts of the Hamiltonian are handled in a wavelet basis. For simplicity, we will discuss only the case where a wave function Ψ is expanded in scaling functions.

$$\Psi(\mathbf{r}) = \sum_{i_1,i_2,i_3} s_{i_1,i_2,i_3} \phi_{i_1,i_2,i_3}(\mathbf{r}) \tag{77}$$

The sum over i_1, i_2, i_3 runs over all the points of a uniform grid. The more general case of adaptive resolution is discussed in the original paper[13].

15.1 Treatment of local potential energy

The local potential $V(\mathbf{r})$ is generally known on the nodes of the uniform grid in the simulation box. Approximating a potential energy matrix element $V_{i,j,k;i',j',k'}$

$$V_{i,j,k;i',j',k'} = \int d\mathbf{r} \phi_{i',j',k'}(\mathbf{r}) V(\mathbf{r}) \phi_{i,j,k}(\mathbf{r})$$

by

$$V_{i,j,k;i',j',k'} \approx \sum_{l,m,n} \phi_{i,j,k}(\mathbf{r}_{l,m,n}) V(\mathbf{r}_{l,m,n}) \phi_{i,j,k}(\mathbf{r}_{l,m,n})$$

gives an extremely slow convergence rate with respect to the number of grid point used to approximate the integral because a single scaling function is not very smooth, i.e. it has a rather low number of continuous derivatives. A. Neelov and S. Goedecker[12] have shown that one should not try to approximate a single matrix element as accurately as possible but that one should try instead to approximate directly the expectation value of the local potential. The reason for this strategy is that the wave function expressed in the Daubechies basis is smoother than a single Daubechies basis function. A single Daubechies scaling function of order 16 has only 4 continuous derivatives. By suitable linear combinations of Daubechies 16 one can however exactly represent polynomials up to degree 7, i.e. functions that have 7 non-vanishing continuous derivatives. The discontinuities get thus canceled by taking suitable linear combinations. Since we use pseudopotentials, our exact wave functions are analytic and they can locally be represented by a Taylor series. We are thus approximating functions that are approximately polynomials of order 7 and the discontinuities cancel to a large degree.

Instead of calculating the exact matrix elements we therefore use matrix elements with respect to a smoothed version $\tilde{\phi}$ of the Daubechies scaling functions.

$$V_{i,j,k;i',j',k'} \approx \sum_{l,m,n} \tilde{\phi}_{i',j',k'}(\mathbf{r}_{l,m,n}) V(\mathbf{r}_{l,m,n}) \tilde{\phi}_{i,j,k}(\mathbf{r}_{l,m,n}) = \sum_{l,m,n} \tilde{\phi}_{0,0,0}(\mathbf{r}_{i'+l,j'+m,k'+n}) V(\mathbf{r}_{l,m,n}) \tilde{\phi}_{0,0,0}(\mathbf{r}_{i+l,j+m,k+n}) \; . \tag{78}$$

The magic filter ω is defined by

$$\omega_{l,m,n} = \tilde{\phi}_{0,0,0}(\mathbf{r}_{l,m,n})$$

The relation between the true functional values, i.e. the scaling function, and ω is shown in figure 14. Even though Eq. 78 is not a particularly good approximation for a single matrix

element it gives an excellent approximation for the expectation values of the local potential energy

$$\int dx \int dy \int dz\ \Psi(x,y,z)V(x,y,z)\Psi(x,y,z)$$

In practice we do not explicitly calculate any matrix elements but we apply only filters to the wave function expansion coefficients as will be shown in the following. This is mathematically equivalent but numerically much more efficient.

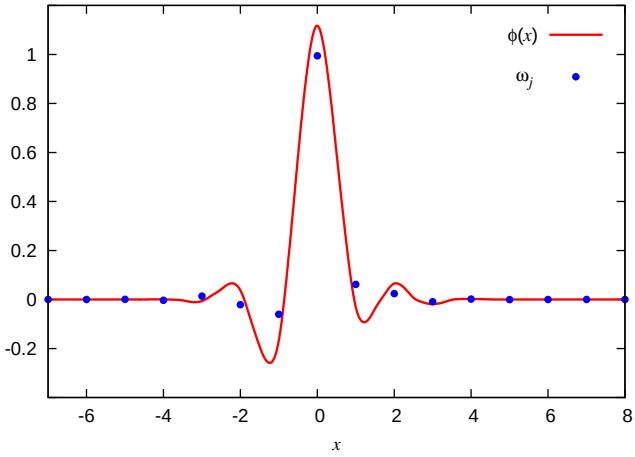


Figure 14. The magic filter ω_i for the least asymmetric Daubechies-16 basis. The values of the magic filter do not coincide with the functional values of the scaling function but represent the behavior of this function in some neighborhood.

Once we have calculated $\tilde{\Psi}_{i,j,k}$ the approximate expectation value ϵ_V of the local potential V for a wave function Ψ is obtained by simple summation on the real space grid:

$$\epsilon_V = \sum_{j_1,j_2,j_3} \tilde{\Psi}_{j_1,j_2,j_3} V_{j_1,j_2,j_3} \tilde{\Psi}_{j_1,j_2,j_3}$$

15.2 Treatment of the non-local pseudopotential

The energy contributions from the non-local pseudopotential have for each angular moment l the form

$$\sum_{i,j} \langle \Psi | p_i \rangle h_{ij} \langle p_j | \Psi \rangle$$

where $|p_i\rangle$ is a pseudopotential projector. When applying the hamiltonian operator on a wave function, such a separable term requires the calculation of

$$|\Psi\rangle \rightarrow |\Psi\rangle + \sum_{i,j} |p_i\rangle h_{ij} \langle p_j | \Psi \rangle \ .$$

It follows from Eq. 51 that the scaling function expansion coefficients for the projectors are given by

$$\int p(\mathbf{r})\phi_{i_1,i_2,i_3}(\mathbf{r})d\mathbf{r} \ . \tag{79}$$

The GTH-HGH pseudopotentials[14,15] have projectors which are written in terms of Gaussian times polynomials. This form of projectors is particularly convenient to be expanded in the Daubechies basis. Since a 3-dimensional Gaussian $\langle \mathbf{r}|\mathbf{p}\rangle = \mathbf{e}^{-\mathbf{cr^2}}\mathbf{x}^{\ell_x}\mathbf{y}^{\ell_y}\mathbf{z}^{\ell_z}$ is a product of three 1-dimensional Gaussians, the 3-dimensional integral 79 can be factorized into a product of three 1-dimensional integrals.

$$\begin{aligned}\int \langle \mathbf{r}|\mathbf{p}\rangle\phi_{\mathbf{i_1},\mathbf{i_2},\mathbf{i_3}}(\mathbf{r})\mathbf{dr} &= W_{i_1}(c,\ell_x)W_{i_2}(c,\ell_y)W_{i_3}(c,\ell_x) \ , \\ W_j(c,\ell) \quad &= \textstyle\int_{-\infty}^{+\infty} e^{-ct^2}t^\ell\phi(t/h-j)dt\end{aligned}$$

The 1-dimensional integrals $W_j(c,\ell)$ are calculated in the following way. We first calculate the scaling function expansion coefficients for scaling functions on a 1-dimensional grid that is 16 times denser. The integration on this dense grid is done by summing the product of the Gaussian and the smoothed scaling function that is obtained by filtering the original scaling function with the magic filter[12]. This integrations scheme based on the magic filter has a convergence rate of h^{14} and we gain therefore a factor of 16^{14} in accuracy by going to a denser grid. This means that the expansion coefficients are for reasonable grid spacings h accurate to machine precision. After having obtained the expansion coefficients with respect to the fine scaling functions we obtain the expansion coefficients with respect to the scaling functions and wavelets on the required resolution level by one-dimensional fast wavelet transformations (Eq. 26). No accuracy with respect to the scaling function coefficients on the lower resolution levels is lost in the wavelet transforms and our representation of the coarse scaling function coefficients of the projectors is therefore typically accurate to nearly machine precision.

16 Final Remarks

Even though wavelet basis sets allow for a very high degree of adaptivity, i.e. many levels of wavelets in Eq. 52, such a high degree of adaptivity causes some numerical overhead that slows down a program. For this reason we have adopted in the BigDFT electronic structure program (http://www-drfmc.cea.fr/sp2m/L_Sim/BigDFT/) only a low degree of adaptivity, namely two resolution levels which are obtained by a set of scaling function augmented by a set of 7 wavelets in the high resolution regions. In most cases, the more rapid variation of the wavefunction around in the chemical bonding region is described by scaling functions plus wavelets whereas the slower variation in the tail regions of the wavefunction is described by scaling functions only. This is typically sufficient since pseudopotentials are used to eliminate the strongly varying core electrons and to account for relativistic effects. All electron wavelet based electronic structure programs do however exist as well[16,17].

References

1. F. Gygi, *Europhys. Lett.* **19**, 617 (1992); Phys. Rev. B **48** 11692 (1993); *Phys. Rev.* B **51** 11190 (1995).
2. Y. Meyer, *"Ondelettes et opérateurs"* Hermann, Paris, 1990.
3. I. Daubechies, *"Ten Lectures on Wavelets"*, SIAM, Philadelphia (1992).
4. S. Goedecker: "Wavelets and their application for the solution of partial differential equations", Presses Polytechniques Universitaires et Romandes, Lausanne, Switzerland 1998, (ISBN 2-88074-3 98-2)
5. W. Sweldens and R. Piessens, *"Wavelets sampling techniques"*, Proceedings of the Joint Statistical Meetings, San Fransisco, August 1993.
6. W. Sweldens, *Appl. Comput. Harmon. Anal.* **3**, 186 (1996).
7. G. Deslauriers and S. Dubuc, *Constr. Approx.* **5**, 49 (1989).
8. C. Tymczak and X. Wang, *Phys. Rev. Lett.* **78**, 3654 (1997).
9. G. Beylkin, *SIAM J. on Numerical Analysis* **6**, 1716 (1992).
10. S. Goedecker, O. Ivanov, *Sol. State Comm.* **105** 665 (1998).
11. Luigi Genovese, Thierry Deutsch, Alexey Neelov, Stefan Goedecker and Gregory Beylkin, J. Chem. Phys. **125**, 074105 (2006)
12. A. I. Neelov and S. Goedecker, J. of. Comp. Phys. **217**, 312-339 (2006)
13. Luigi Genovese, Alexey Neelov, Stefan Goedecker, Thierry Deutsch, Alireza Ghasemi, Oded Zilberberg, Anders Bergman, Mark Rayson and Reinhold Schneider, J. Chem. Phys. **129**, 014109 (2008)
14. S. Goedecker, M. Teter, and J. Hutter, *Phys. Rev. B* **54**, 1703, (1996).
15. C. Hartwigsen, S. Goedecker and J. Hutter, Phys. Rev. B **58**, 3641 (1998)
16. T.D. Engeness and T.A. Arias, Phys. Rev. B65, 165106-1 (2002)
17. Robert J. Harrison, George I. Fann, Takeshi Yanai, Zhengting Gan, and Gregory Beylkin J. Chem. Phys. **121** 11587 (2004)

Introduction to Parallel Computing

Bernd Mohr

Institute for Advanced Simulation (IAS)
Jülich Supercomputing Centre (JSC)
Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: b.mohr@fz-juelich.de

The major parallel programming models for scalable parallel architectures are the message passing model and the shared memory model. This article outlines the main concepts of these models as well as the industry standard programming interfaces MPI and OpenMP. To exploit the potential performance of parallel computers, programs need to be carefully designed and tuned. We will discuss design decisions for good performance as well as programming tools that help the programmer in program tuning.

1 Introduction

Many applications like numerical simulations in industry and research as well as commercial applications such as query processing, data mining, and multi-media applications require more compute power than provided by sequential computers. Current hardware architectures offering high performance do not only exploit parallelism within a single processor via multiple CPU cores but also apply a medium to large number of processors concurrently to a single computation. High-end parallel computers currently (2009) deliver up to 1 Petaflop/s (10^{15} floating point operations per second) and are developed and exploited within the ASC (Advanced Simulation and Computing) program of the Department of Energy in the USA and PRACE (Partnership for Advanced Computing in Europe) in Europe. In addition, the current trend to multi-core processors also requires parallel programming to fully exploit the compute power of the multiple cores.

This article concentrates on programming numerical applications on parallel computer architectures introduced in Section 1.1. Parallelization of those applications centers around selecting a decomposition of the data domain onto the processors such that the workload is well balanced and the communication between processors is reduced (Section 1.2)[4].

The parallel implementation is then based on either the message passing or the shared memory model (Section 2). The standard programming interface for the message passing model is MPI (Message Passing Interface)[8–12], offering a complete set of communication routines (Section 3). OpenMP[13–15] is the standard for directive-based shared memory programming and will be introduced in Section 4.

Since parallel programs exploit multiple threads of control, debugging is even more complicated than for sequential programs. Section 5 outlines the main concepts of parallel debuggers and presents TotalView[21] and DDT[3], the most widely available debuggers for parallel programs.

Although the domain decomposition is key to good performance on parallel architectures, program efficiency also heavily depends on the implementation of the communication and synchronization required by the parallel algorithms and the implementation techniques chosen for sequential kernels. Optimizing those aspects is very system dependent and thus, an interactive tuning process consisting of measuring performance data and

applying optimizations follows the initial coding of the application. The tuning process is supported by programming model specific performance analysis tools. Section 6 presents basic performance analysis techniques.

1.1 Parallel Architectures

A *parallel computer* or *multi-processor system* is a computer utilizing more than one processor. A common way to classify parallel computers is to distinguish them by the way how processors can access the system's main memory because this influences heavily the usage and programming of the system.

In a *distributed memory architecture* the system is composed out of single-processor nodes with local memory. The most important characteristic of this architecture is that access to the local memory is faster than to remote memory. It is the challenge for the programmer to assign data to the processors such that most of the data accessed during the computation are already in the node's local memory. Two major classes of distributed memory computers can be distinguished:

No Remote Memory Access (NORMA) computers do not have any special hardware support to access another node's local memory directly. The nodes are only connected through a computer network. Processors obtain data from remote memory only by exchanging messages over this network between processes on the requesting and the supplying node. Computers in this class are sometimes also called **Network Of Workstations (NOW)** or **Clusters Of Workstations (COW)**.

Remote Memory Access (RMA) computers allow to access remote memory via specialized operations implemented by hardware, however the hardware does not provide a global address space, i.e., a memory location is not determined via an address in a shared linear address space but via a tuple consisting of the processor number and the local address in the target processor's address space.

The major advantage of distributed memory systems is their ability to scale to a very large number of nodes. Today (2009), systems with more than 210,000 cores have been built. The disadvantage is that such systems are very hard to program.

In contrast, a *shared memory architecture* provides (in hardware) a global address space, i.e., all memory locations can be accessed via usual load and store operations. Access to a remote location results in a copy of the appropriate cache line in the processor's cache. Therefore, such a system is much easier to program. However, shared memory systems can only be scaled to moderate numbers of processors, typically 64 or 128. Shared memory systems are further classified according to the quality of the memory accesses:

Uniform Memory Access (UMA) computer systems feature one global shared memory subsystem which is connected to the processors through a central bus or memory switch. All of the memory is accessible to all processors in the same way. Such a system is also often called **Symmetrical Multi Processor (SMP)**.

Non Uniform Memory Access (NUMA) computers are more scalable by physically distributing the memory but still providing a hardware implemented global address space. Therefore access to memory local or close to a processor is faster than to remote memory. If such a system has additional hardware which also ensures that multiple copies of data stored in different cache lines of the processors is kept coherent, i.e.,

the copies always do have the same value, then it is called a **Cache-Coherent Non Uniform Memory Access (ccNUMA)** system. ccNUMA systems offer the abstraction of a shared linear address space resembling physically shared memory systems. This abstraction simplifies the task of program development but does not necessarily facilitate program tuning.

While most of the early parallel computers were simple single processor NORMA systems, today's large parallel systems are typically *hybrid systems*, i.e., shared memory NUMA nodes with a moderate number of processors are connected together to form a distributed memory cluster system.

1.2 Data Parallel Programming

Applications that scale to a large number of processors usually perform computations on large data domains. For example, crash simulations are based on partial differential equations that are solved on a large finite element grid and molecular dynamics applications simulate the behavior of a large number of particles. Other parallel applications apply linear algebra operations to large vectors and matrices. The elemental operations on each object in the data domain can be executed in parallel by the available processors.

The scheduling of operations to processors is determined by a *domain decomposition*[5] specified by the programmer. Processors execute those operations that determine new values for elements stored in local memory (owner-computes rule). While processors execute an operation, they may need values from other processors. The domain decomposition has thus to be chosen so that the distribution of operations is balanced and the communication is minimized. The third goal is to optimize single node computation, i.e., to be able to exploit the processor's pipelines and the processor's caches efficiently.

A good example for the design decisions taken when selecting a domain decomposition is Gaussian elimination[1]. The main structure of the matrix during the steps of the algorithm is outlined in Figure 1.

The goal of this algorithm is to eliminate all entries in the matrix below the main diagonal. It starts at the top diagonal element and subtracts multiples of the first row from the second and subsequent rows to end up with zeros in the first column. This operation is repeated for all the rows. In later stages of the algorithm the actual computations have to be done on rectangular sections of decreasing size. If the main diagonal element of the current row is zero, a pivot operation has to be performed. The subsequent row with the maximum value in this column is selected and exchanged with the current row.

A possible distribution of the matrix is to decompose its columns into blocks, one block for each processor. The elimination of the entries in the lower triangle can then be performed in parallel where each processor computes new values for its columns only. The main disadvantage of this distribution is that in later computations of the algorithm only a subgroup of the processors is actually doing any useful work since the computed rectangle is getting smaller.

To improve load balancing, a cyclic column distribution can be applied. The computations in each step of the algorithm executed by the processors differ only in one column.

In addition to load balancing also communication needs to be minimized. Communication occurs in this algorithm for broadcasting the current column to all the processors since it is needed to compute the multiplication factor for the row. If the domain decomposition

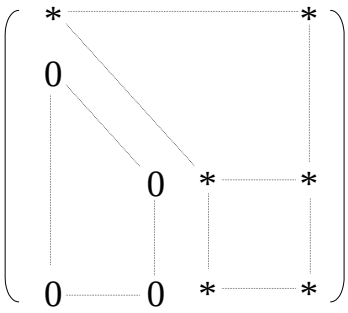


Figure 1. Structure of the matrix during Gaussian elimination.

is a row distribution, which eliminates the need to communicate the current column, the current row needs to be broadcast to the other processors.

If we consider also the pivot operation, communication is necessary to select the best row when a row-wise distribution is applied since the computation of the global maximum in that column requires a comparison of all values.

Selecting the best domain decomposition is further complicated due to optimizing single node performance. In this example, it is advantageous to apply BLAS3[2] operations for the local computations. These operations make use of blocks of rows to improve cache utilization. Blocks of rows can only be obtained if a block-cyclic distribution is applied, i.e., columns are not distributed individually but blocks of columns are cyclically distributed.

This discussion makes clear, that choosing a domain decomposition is a very complicated step in program development. It requires deep knowledge of the algorithm's data access patterns as well as the ability to predict the resulting communication.

2 Programming Models

Programming parallel computers is almost always done via the so-called *Single Program Multiple Data* (SPMD) model. SPMD means that the same program (executable code) is executed on all processors taking part in the computation, but it computes on different parts of the data which were distributed over the processors based on a specific domain decomposition. If computations are only allowed on specific processors, this has to be explicitly programmed by using conditional programming constructs (e.g., with `if` or `where` statements). There are two main programming models, *message passing* and *shared memory*, offering different features for implementing applications parallelized by domain decomposition.

The message passing model is based on a set of processes with private data structures. Processes communicate by exchanging messages with special send and receive operations. It is a natural fit for programming distributed memory machines but also can be used on shared memory computers. The domain decomposition is implemented by developing a code describing the local computations and local data structures of a single process. Thus, global arrays have to be split up and only the local part has to be allocated in a process. This handling of global data structures is called *data distribution*. Computations on the global arrays also have to be transformed, e.g., by adapting the loop bounds, to ensure that

only local array elements are computed. Access to remote elements has to be implemented via explicit communication, temporary variables have to be allocated, messages have to be constructed and transmitted to the target process.

The shared memory model is based on a set of threads that is created when parallel operations are executed. This type of computation is also called *fork-join parallelism*. Threads share a global address space and thus access array elements via a global index. The main parallel operations are *parallel loops* and *parallel sections*. Parallel loops are executed by a set of threads also called a *team*. The iterations are distributed among the threads according to a predefined strategy. This scheduling strategy implements the chosen domain decomposition. Parallel sections are also executed by a team of threads but the tasks assigned to the threads implement different operations. This feature can for example be applied if domain decomposition itself does not generate enough parallelism and whole operations can be executed in parallel since they access different data structures.

In the shared memory model, the distribution of data structures onto the node memories is not enforced by decomposing global arrays into local arrays, but the global address space is distributed onto the memories by the operating system. For example, the pages of the virtual address space can be distributed cyclically or can be assigned at first touch. The chosen domain decomposition thus has to take into account the granularity of the distribution, i.e., the size of pages, as well as the system-dependent allocation strategy.

While the domain decomposition has to be hard-coded into the message passing program, it can easily be changed in a shared memory program by selecting a different scheduling strategy for parallel loops.

Another advantage of the shared memory model is that automatic and incremental parallelization is supported. While automatic parallelization leads to a first working parallel program, its efficiency typically needs to be improved. The reason for this is that parallelization techniques work on a loop-by-loop basis and do not globally optimize the parallel code via a domain decomposition. In addition, dependence analysis, the prerequisite for automatic parallelization, is limited to access patterns known at compile time. The biggest disadvantage of this model is that it can only be used on shared memory computers.

In the shared memory model, a first parallel version is relatively easy to implement and can be incrementally tuned. In the message passing model instead, the program can be tested only after finishing the full implementation. Subsequent tuning by adapting the domain decomposition is usually time consuming.

3 MPI

The Message Passing Interface (MPI)[8–12] was mainly developed between 1993 and 1997. It is a community standard which standardizes the calling interface for a communication and synchronization function library. It provides Fortran 77, Fortran 90, C and C++ language bindings. It includes routines for point-to-point communication, collective communication, one-sided communication, parallel IO, and dynamic task creation. Currently, almost all available open-source and commercial MPI implementations support the 2.0 standard with the exception of dynamic task creation, which is only implemented by a few. In 2008, an update and clarification of the standard was published as Version 2.1 and work has begun to define further enhancements (version 3.x). For a simple example see the appendix.

3.1 MPI Basic Routines

MPI consists of more than 320 functions. But realistic programs can already be developed based on no more than six functions:

MPI_Init initializes the library. It has to be called at the beginning of a parallel operation before any other MPI routines are executed.

MPI_Finalize frees any resources used by the library and has to be called at the end of the program.

MPI_Comm_size determines the number of processors executing the parallel program.

MPI_Comm_rank returns the unique process identifier.

MPI_Send transfers a message to a target process. This operation is a blocking send operation, i.e., it terminates when the message buffer can be reused either because the message was copied to a system buffer by the library or because the message was delivered to the target process.

MPI_Recv receives a message. This routine terminates if a message was copied into the receive buffer.

3.2 MPI Communicator

All communication routines depend on the concept of a *communicator*. A communicator consists of a process group and a communication context. The processes in the process group are numbered from zero to process count - 1. The process number returned by MPI_Comm_rank is the identification in the process group of the communicator which is passed as a parameter to this routine.

The communication context of the communicator is important in identifying messages. Each message has an integer number called a *tag* which has to match a given selector in the corresponding receive operation. The selector depends on the communicator and thus on the communication context. It selects only messages with a fitting tag and having been sent relative to the same communicator. This feature is very useful in building parallel libraries since messages sent inside the library will not interfere with messages outside if a special communicator is used in the library. The default communicator that includes all processes of the application is MPI_COMM_WORLD.

3.3 MPI Collective Operations

Another important class of operations are *collective operations*. Collective operations are executed by a process group identified via a communicator. All the processes in the group have to perform the same operation. Typical examples for such operations are:

MPI_Barrier synchronizes all processes. None of the processes can proceed beyond the barrier until all the processes started execution of that routine.

MPI_Bcast allows to distribute the same data from one process, the so-called *root* process, to all other processes in the process group.

MPI_Scatter also distributes data from a root process to a whole process group, but each receiving process gets different data.

MPI_Gather collects data from a group of processes at a root process.

MPI_Reduce performs a global operation on the data of each process in the process group. For example, the sum of all values of a distributed array can be computed by first summing up all local values in each process and then summing up the local sums to get a global sum. The latter step can be performed by the reduction operation with the parameter MPI_SUM. The result is delivered to a single target processor.

3.4 MPI IO

Data parallel applications make use of the IO subsystem to read and write big data sets. These data sets result from replicated or distributed arrays. The reasons for IO are to read input data, to pass information to other programs, e.g., for visualization, or to store the state of the computation to be able to restart the computation in case of a system failure or if the computation has to be split into multiple runs due to its resource requirements.

IO can be implemented in three ways:

1. Sequential IO

   A single node is responsible to perform the IO. It gathers information from the other nodes and writes it to disk or reads information from disk and scatters it to the appropriate nodes. Whereas this approach might be feasible for small amounts of data, it bears serious scalability issues, as modern IO subsystems can only be utilized efficiently with parallel data streams and aggregated waiting time increases rapidly at larger scales.

2. Private IO

   Each node accesses its own files. The big advantage of this implementation is that no synchronization among the nodes is required and very high performance can be obtained. The major disadvantage is that the user has to handle a large number of files. For input the original data set has to be split according to the distribution of the data structure and for output the process-specific files have to be merged into a global file for post-processing.

3. Parallel IO

   In this implementation all the processes access the same file. They read and write only those parts of the file with relevant data. The main advantages are that no individual files need to be handled and that reasonable performance can be reached. The parallel IO interface of MPI provides flexible and high-level means to implement applications with parallel IO.

Files accessed via MPI IO routines have to be opened and closed by collective operations. The open routine allows to specify hints to optimize the performance such as whether the application might profit from combining small IO requests from different nodes, what size is recommended for the combined request, and how many nodes should be engaged in merging the requests.

The central concept in accessing the files is the *view*. A view is defined for each process and specifies a sequence of data elements to be ignored and data elements to be read or written by the process. When reading or writing a distributed array the local information can be described easily as such a repeating pattern. The IO operations read and write

a number of data elements on the basis of the defined view, i.e., they access the local information only. Since the views are defined via runtime routines prior to the access, the information can be exploited in the library to optimize IO.

MPI IO provides blocking as well as nonblocking operations. In contrast to blocking operations, the nonblocking ones only start IO and terminate immediately. If the program depends on the successful completion of the IO it has to check it via a test function. Besides the collective IO routines which allow to combine individual requests, also non-collective routines are available to access shared files.

3.5 MPI Remote Memory Access

Remote memory access (RMA) operations (also called *one-sided communication*) allow to access the address space of other processes without participation of the other process. The implementation of this concept can either be in hardware, such as in the CRAY T3E, or in software via additional threads waiting for requests. The advantages of these operations are that the protocol overhead is much lower than for normal send and receive operations and that no polling or global communication is required for setting up communication.

In contrast to explicit message passing where synchronization happens implicitly, accesses via RMA operations need to be protected by explicit synchronization operations.

RMA communication in MPI is based on the *window concept*. Each process has to execute a collective routine that defines a window, i.e., the part of its address space that can be accessed by other processes.

The actual access is performed via *put* and *get* operations. The address is defined by the target process number and the displacement relative to the starting address of the window for that process.

MPI also provides special synchronization operations relative to a window. The MPI_Win_fence operation synchronizes all processes that make some address ranges accessible to other processes. It is a collective operation that ensures that all RMA operations started before the fence operation terminate before the target process executes the fence operation and that all RMA operations of a process executed after the fence operation are executed after the target process executed the fence operation. There are also more fine grained synchronization methods available in the form of the General Active Target Synchronization or via locks.

4 OpenMP

OpenMP[13–15] is a directive-based programming interface for the shared memory programming model. It consists of a set of directives and runtime routines for Fortran 77 (published 1997), for Fortran 90 (2000), and a corresponding set of pragmas for C and C++ (1998). In 2005, a combined Fortran, C, and C++ standard (Version 2.5) and 2008, an update (Version 3.0) were published.

Directives are special comments that are interpreted by the compiler. Directives have the advantage that the code is still a sequential code that can be executed on sequential machines (by ignoring the directives/pragmas) and therefore there is no need to maintain separate sequential and parallel versions.

Directives start and terminate parallel regions. When the master thread hits a parallel region a team of threads is created or activated. The threads execute the code in parallel and are synchronized at the beginning and the end of the computation. After the final synchronization the master thread continues sequential execution after the parallel region. The main directives are:

!$OMP PARALLEL DO specifies a loop that can be executed in parallel. The DO loop's iterations can be distributed among the set of threads according to various scheduling strategies including STATIC(CHUNK), DYNAMIC(CHUNK), and GUIDED(CHUNK). STATIC(CHUNK) distribution means that the set of iterations are consecutively distributed among the threads in blocks of CHUNK size (resulting in block and cyclic distributions). DYNAMIC(CHUNK) distribution implies that iterations are distributed in blocks of CHUNK size to threads on a first-come-first-served basis. GUIDED (CHUNK) means that blocks of exponentially decreasing size are assigned on a first-come-first-served basis. The size of the smallest block is determined by CHUNK size.

!$OMP PARALLEL SECTIONS starts a set of sections that are each executed in parallel by a team of threads.

!$OMP PARALLEL introduces a code region that is executed redundantly by the threads. It has to be used very carefully since assignments to global variables will lead to conflicts among the threads and possibly to nondeterministic behavior.

!$OMP DO / FOR is a work sharing construct and may be used within a parallel region. All the threads executing the parallel region have to cooperate in the execution of the parallel loop. There is no implicit synchronization at the beginning of the loop but a synchronization at the end. After the final synchronization all threads continue after the loop in the replicated execution of the program code.

The main advantage of this approach is that the overhead for starting up the threads is eliminated. The team of threads exists during the execution of the parallel region and need not be built before each parallel loop.

!$OMP SECTIONS is also a work sharing construct that allows the current team of threads executing the surrounding parallel region to cooperate in the execution of the parallel sections.

!$OMP TASK is only available with the new version 3.0 of the standard and greatly simplifies the parallelization on non-loop constructs by allowing to dynamically specify portions of the programs which can run independently.

Program data can either be shared or private. While threads do have their own copy of private data, only one copy exists of shared data. This copy can be accessed by all threads. To ensure program correctness, OpenMP provides special synchronization constructs. The main constructs are *barrier synchronization* enforcing that all threads have reached this synchronization operation before execution continues and *critical sections*. Critical sections ensure that only a single thread can enter the section and thus, data accesses in such a section are protected from race conditions. For example, a common situation for a critical section is the accumulation of values. Since an accumulation consists of a read and a write operation unexpected results can occur if both operations are not surrounded by a critical section. For a simple example of an OpenMP parallelization see the appendix.

5 Parallel Debugging

Debugging parallel programs is more difficult than debugging sequential programs not only since multiple processes or threads need to be taken into account but also because program behavior might not be deterministic and might not be reproducible. These problems are not solved by current state-of-the-art commercial parallel debuggers. They only deal with the first problem by providing menus, displays, and commands that allow to inspect individual processes and execute commands on individual or all processes.

Two widely used debuggers are TotalView from Totalview Technologies[21] and DDT from Allinea[3]. They provide breakpoint definition, single stepping, and variable inspection for parallel programs via an interactive interface. The programmer can execute those operations for individual processes and groups of processes. They also provides some means to summarize information such that equal information from multiple processes is combined into a single information and not repeated redundantly. They also support MPI and OpenMP programs on many platforms.

6 Parallel Performance Analysis

Performance analysis is an iterative subtask during program development. The goal is to identify program regions that do not perform well. Performance analysis is structured into three phases:

1. Measurement

   Performance analysis is done based on information on runtime events gathered during program execution. The basic events are, for example, cache misses, termination of a floating point operation, start and stop of a subroutine or message passing operation. The information on individual events can be summarized during program execution (*profiling*) or individual trace records can be collected for each event (*tracing*).

2. Analysis

   During analysis the collected runtime data are inspected to detect *performance problems*. Performance problems are based on *performance properties*, such as the existence of message passing in a program region, which have a condition for identifying it and a severity function that specifies its importance for program performance.

   Current tools support the user in checking the conditions and the severity by a visualization of the program behavior. Future tools might be able to automatically detect performance properties based on a specification of possible properties. During analysis the programmer applies a threshold. Only performance properties whose severity exceeds this threshold are considered to be performance problems.

3. Ranking

   During program analysis the severest performance problems need to be identified. This means that the problems need to be ranked according to the severity. The most severe problem is called the *program bottleneck*. This is the problem the programmer tries to resolve by applying appropriate program transformations.

Current techniques for performance data collection are *profiling* and *tracing*. Profiling collects summary data only. This can be done via *sampling*. The program is regularly interrupted, e.g., every 10 ms, and the information is added up for the source code location which was executed in this moment. For example, the UNIX profiling tool *prof* applies this technique to determine the fraction of the execution time spent in individual subroutines.

A more precise profiling technique is based on *instrumentation*, i.e., special calls to a *monitoring library* are inserted into the program. This can either be done in the source code by the compiler or specialized tools, or can be done in the object code. While the first approach allows to instrument more types of regions, for example, loops and vector statements, the latter allows to measure data for programs where no source code is available. The monitoring library collects the information and adds it to special counters for the specific region.

Tracing is a technique that collects information for each event. This results, for example, in very detailed information for each instance of a subroutine and for each message sent to another process. The information is stored in specialized trace records for each event type. For example, for each start of a send operation, the time stamp, the message size and the target process can be recorded, while for the end of the operation, the time stamp and bandwidth are stored.

The trace records are stored in the memory of each process and are written to a trace file either when the buffer is filled up or when the program terminates. The individual trace files of the processes are merged together into one trace file ordered according to the time stamps of the events.

Profiling has the advantage to be of moderate size while trace information tends to be very large. The disadvantage of profiling is that it is not fine grained; the behavior of individual instances of subroutines can for example not be investigated since all the information has been summed up.

Widely used performance tools include TAU[19,20] from the University of Oregon, Vampir[22,23] from the Technical University of Dresden, and Scalasca[17,18] from the Jülich Supercomputing Centre.

7 Summary

This article gave an overview of parallel programming models as well as programming tools. Parallel programming will always be a challenge for programmers. Higher-level programming models and appropriate programming tools only facilitate the process but do not make it a simple task.

While programming in MPI offers the greatest potential performance, shared memory programming with OpenMP is much more comfortable due to the global style of the resulting program. The sequential control flow among the parallel loops and regions matches much better with the sequential programming model all the programmers are trained for.

Although programming tools were developed over years, the current situation seems not to be very satisfying. Program debugging is done per thread, a technique that does not scale to larger numbers of processors. Performance analysis tools do also suffer scalability limitations and, in addition, the tools are complicated to use. The programmers have to be experts for performance analysis to understand potential performance problems, their proof conditions, and their severity. In addition they have to be experts for powerful but

also complex user interfaces.

Future research in this area has to try to automate performance analysis tools, such that frequently occurring performance problems can be identified automatically. First automatic tools are already available: ParaDyn[7] from the University of Wisconsin-Madison and KOJAK[6]/Scalasca[17,18] from the Jülich Supercomputing Centre.

A second important trend that will effect parallel programming in the future is the move towards clustered shared memory systems with nodes consisting of multi-core processors. This introduces a potentially 3-level parallelism hierarchy (machine - node - processor). Clearly, a hybrid programming approach will be applied on those systems for best performance, combining message passing between the individual SMP nodes and shared memory programming in a node. This programming model will lead to even more complex programs and program development tools have to be enhanced to be able to help the user in developing these codes.

A promising approach to reduce complexity in parallel programming in the future are so-called *partitioned global address space* (PGAS) languages[16], such as Unified Parallel C (UPC) or Co-array Fortran (CAF) which provide simple means to distribute data and communicate implicitly via efficient one-sided communication.

Appendix

This appendix provides three versions of a simple example of a scientific computation. It computes the value of π by numerical integration:

$$\pi = \int_0^1 f(x)dx \quad \text{with} \quad f(x) = \frac{4}{1+x^2}$$

This integral can be approximated numerically by the midpoint rule:

$$\pi \approx \frac{1}{n}\int_1^n f(x_i) \quad \text{with} \quad x_i = \frac{(i-0.5)}{n} \quad \text{for} \quad i = 1,\ldots,n$$

Larger values of the parameter n will give us more accurate approximations of π. This is not, in fact, a very good way to compute π, but it makes a good example because it has the typical, complete structure of a numerical simulation program (initialization - loop-based calculation - wrap-up), and the whole source code fits one one page or slide.

To parallelize the example, each process/thread computes and adds up the areas for a different subset of the rectangles. At the end of the computation, all of the local sums are combined into a global sum representing the value of π.

MPI Version of Example Program

The following listing shows a Fortran90 implementation of the π numerical integration example parallelized with the help of MPI.

```
1  program pi_mpi
2  implicit none
3  include 'mpif.h'
4  integer          :: i, n, ierr, myrank, numprocs
5  double precision :: f, x, sum, pi, h, mypi
6
7  call MPI_Init(ierr)
8  call MPI_Comm_rank(MPI_COMM_WORLD, myrank, ierr)
9  call MPI_Comm_size(MPI_COMM_WORLD, numprocs, ierr)
10
11 if ( myrank == 0 ) then
12    write(*,*) "number of intervals?"
13    read(*,*) n
14 end if
15
16 call MPI_Bcast(n, 1, MPI_INTEGER, 0, MPI_COMM_WORLD, ierr)
17
18 h = 1.0d0 / n
19 sum = 0.0d0
20 do i = myrank+1, n, numprocs
21    x = (i - 0.5d0) * h
22    sum = sum + (4.d0/(1.d0 + x*x))
23 end do
24 mypi = h * sum
25
26 call MPI_Reduce(mypi, pi, 1, MPI_DOUBLE_PRECISION, &
27                 MPI_SUM, 0, MPI_COMM_WORLD, ierr)
28
29 if ( myrank == 0 ) then
30   write(*, fmt="(A, F16.12)") "Value of pi is ", pi
31 endif
32
33 call MPI_Finalize(ierr)
34 end program
```

First, the MPI system has to be initialized (lines 7 to 9) and terminated (line 33) with the necessary MPI calls. Next, the input of parameters (line 11 to 14) and the output of results (lines 29 to 31) has to be restricted so that it is only executed by one processor. Then, the input has to be broadcasted to the other processors (line 16). The biggest (and most complicated) change is to program the distribution of work and data. The do-loop in line 20 has to be changed so that each processor only calculates and summarizes its part of the distributed computations. Finally, the reduce call in lines 26/27 collects the local sums and delivers the global sum to processor 0.

Sequential and OpenMP Version of Example Program

The following listing shows the corresponding implementation of the π integration example using OpenMP. As one can see, because of the need to explicitly program all aspects of the parallelization, the MPI version is almost twice as long as the OpenMP version. Although this is clearly more work, it gives a programmer much more ways to express and control parallelism. Also, the MPI version will run on all kinds of parallel computers, while OpenMP is restricted to the shared memory architecture.

As OpenMP is based on directives (which are plain comments in a non-OpenMP compilation mode), it is at the same time also a sequential implementation of the example.

```
1  program pi_omp
2  implicit none
3  integer          :: i, n
4  double precision :: f, x, sum, pi, h
5
6  write(*,*) "number of intervals?"
7  read(*,*) n
8
9  h = 1.0d0 / n
10 sum = 0.0d0
11 !$omp parallel do private(i,x) reduction(+:sum)
12 do i = 1, n
13    x = (i - 0.5d0) * h
14    sum = sum + (4.d0/(1.d0 + x*x))
15 end do
16 pi = h * sum
17
18 write(*, fmt="(A, F16.12)") "Value of pi is ", pi
19 end program
```

The OpenMP directive in line 11 declares the following do-loop as parallel resulting in a concurrent execution of loop iterations. As the variables `i` and `x` are used to store values during the execution of the loop, they have to be declared private, so that each thread executing iterations has its own copy. The variable `h` is only read, so it can be shared. Finally, it is specified that there is a reduction (using addition) over the variable `sum`.

References

1. D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall (1989).
2. J. J. Dongarra, J. Du Croz, I. S. Duff, and S. Hammarling, A set of Level 3 Basic Linear Algebra Subprograms, *ACM Trans. Math. Soft.*, 16:1–17, (1990).
3. Allinea: *DDT*, `http://allinea.com/`.
4. I. Foster, *Designing and Building Parallel Programs*, Addison Wesley (1994).

5. G. Fox, *Domain Decomposition in Distributed and Shared Memory Environments*, International Conference on Supercomputing June 8-12, 1987, Athens, Greece, Lecture Notes in Computer Science 297, edited by C. Polychronopoulos (1987).
6. F. Wolf and B. Mohr, Automatic Performance Analysis of Hybrid MPI/OpenMP Applications. *Journal of Systems Architecture*, 49(10–11):421–439 (2003).
7. B. P. Miller, M. D. Callaghan, J. M. Cargille, J. K. Hollingsworth, R. B. Irvine, K. L. Karavanic, K. Kunchithapadam, and T. Newhall, The Paradyn Parallel Performance Measurement Tool, *IEEE Computer*, Vol. 28, No. 11, 37–46 (1995).
8. MPI Forum: *Message Passing Interface*, `http://www.mpi-forum.org/`.
9. M. Snir, S. Otto, S. Huss-Lederman, D. Walker, and J. Dongarra, *MPI - the Complete Reference, Volume 1, The MPI Core*, 2nd ed., MIT Press (1998).
10. W. Gropp, S. Huss-Lederman, A. Lumsdaine, E. Lusk, B. Nitzberg, W. Saphir, and M. Snir, *MPI - the Complete Reference, Volume 2, The MPI Extensions*, MIT Press (1998).
11. W. Gropp, E. Lusk, A. , Skjellum, *Using MPI, 2nd Edition*, MIT Press (1999).
12. W. Gropp, E. Lusk, R. Thakur, *Using MPI-2: Advanced Features of the Message Passing Interface*, MIT Press (1999).
13. OpenMP Forum: *OpenMP Standard*, `http://www.openmp.org/`.
14. L. Dagum and R. Menon, *OpenMP: An Industry-Standard API for Shared-memory Programming*, IEEE Computational Science & Engineering, Vol. 5, No. 1, 46–55 (1998).
15. B. Chapman, G. Jost, R. van der Pas, *Using OpenMP: Portable Shared Memory Parallel Programming*, MIT Press (2007).
16. C. Coarfa, Y. Dotsenko, J. Mellor-Crummey, F. Cantonnet, T. El-Ghazawi, A. Mohanty, Y. Yao, An Evaluation of Global Address Space Languages: Co-Array Fortran and Unified Parallel C, *Proceedings of the ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPOPP 2005)*, ACM, (2005)
17. M. Geimer, F. Wolf, B. J. N. Wylie, B. Mohr, Scalable parallel trace-based performance analysis. *Proc. 13th European PVM/MPI Users' Group Meeting*, Bonn, Germany. Volume 4192 of LNCS, Springer, pp. 303–312, (2006).
18. B. J. N. Wylie, M. Geimer, F. Wolf, Performance measurement and analysis of large-scale parallel applications on leadership computing systems, *Scientific Programming*, 16(2–3), Special Issue on Large-Scale Programming Tools and Environments, pp. 167–181, (2008).
19. A. Malony, S. Shende, A. Morris, Performance technology for productive parallel computing, *Advanced Scientific Computing Research Computer Science FY2005 Accomplishments*, U.S. Dept. of Energy Office of Science, (2005).
20. S. Shende, A. Malony, The TAU parallel performance system, *International Journal of High Performance Computing Applications*, 20, pp. 287–331, SAGE Publications, (2006).
21. Totalview Technologies: *TotalView*, `http://www.totalviewtech.com/`.
22. H. Brunst, W. E. Nagel, Scalable performance analysis of parallel systems: Concepts and experiences. *Proc. of the Parallel Computing Conference (ParCo)*, Dresden, Germany, pp. 737–744, (2003).
23. H. Brunst, *Integrative Concepts for Scalable Distributed Performance Analysis and Visualization of Parallel Programs*, PhD thesis, TU Dresden, Shaker, Aachen, (2008).

Strategies for Implementing Scientific Applications on Parallel Computers

Bernd Körfgen and Inge Gutheil

Institute for Advanced Simulation (IAS)
Jülich Supercomputing Centre (JSC)
Forschungszentrum Jülich
52425 Jülich, Germany
E-mail: {B.Koerfgen, I.Gutheil}@fz-juelich.de

This contribution presents two examples for the numerical treatment of partial differential equations using parallel algorithms / computers. The first example solves the Poisson equation in two dimensions; the second partial differential equation describes the physical process of vibration of a membrane. Both problems are tackled with different strategies: The Poisson equation is solved by means of the simple Jacobi algorithm and a suitable parallelization scheme is discussed; in the second case the parallel calculation is performed with the help of the ScaLAPACK library and the issue of data distribution is addressed. Finally benchmarks of linear algebra libraries on the IBM Blue Gene/P architecture and for the PowerXCell processor are shown.

1 Motivation

The growth of processing power of computer systems observed through the last decades allowed scientists and engineers to perform more and more complex and realistic simulations. The driving force for this development was the exponentially increasing density of integrated circuits, e.g. processors, which can be empirically described by Moore's law. The key point of Moore's law is the observation that the circuit density of electronic devices doubles approximately every two years. The growing number of components together with the accompanying higher clock frequencies permitted to perform more intricate computations in shorter times.

Today this growth of circuit density and clock frequencies becomes more and more difficult to achieve. Furthermore the demand for compute power grows even faster (improved spatial / time resolution of models or the introduction of additional interactions / effects are required) and forces the developers of supercomputers to find new strategies to increase the available compute power. One approach which has become quite popular over the last two decades is the utilization of parallel computers. Many of nowadays parallel architectures use multi-core processors as building blocks in order to obtain the necessary compute power. Adding up ever more of these components generates the problem of excessive power consumption; not only to supply the electrical power but likewise to handle the produced heat are real challenges for the design of supercomputers.

Possible strategies to overcome the so-called 'power wall' are the reduction of clock frequencies and thus power consumption and the usage of special high performance processors which do more computations per clock cycle. The first concept has been realized with IBM's highly scalable **Blue Gene**[1] architecture. The latter strategy has been implemented by means of the **Cell processor**[2] which is together with Opteron processors the workhorse of the IBM Roadrunner[3] - the first supercomputer to reach **one Petaflop/s** (10^{15}

floating point operations per second). The price one has to pay for this development is the growing complexity of these heterogeneous systems.

The development of scientific applications on parallel computers, especially on highly scalable heterogeneous platforms poses a challenge to every scientist and engineer. The difficulties arising are to some extent problem specific and have to be resolved as the case arises assisted where possible by **parallel performance tools** like Scalasca[4]. Other tasks are generic and the application developer can resort to well established algorithms or even ready to use software solutions. Examples for these generic methods are **graph partitioning** algorithms for the decomposition of the problem (parallel load balancing) or **linear algebra** algorithms.

We will focus in the following on linear algebra because these algorithms are the core of many simulation codes and thus determine the efficiency and scalability of the whole application.

2 Linear Algebra

Numerical linear algebra is an active field of research which provided over the years many methods / algorithms for the treatment of standard problems like the solution of systems of linear equations, the factorization of matrices, the calculation of eigenvalues / eigenvectors etc.[5]. The most suitable algorithm for a given linear algebra problem, e.g. arising in a scientific application, has to be determined depending on the properties of the system / matrix (see for instance Ref. 6) like:

- **symmetry**
- **definiteness** (positive, negative, . . .)
- **non-zero structure** (dense, sparse, banded)
- **real** or **complex** coefficients

and so on. Furthermore the scientist has to decide whether to use a **direct** solver, leading to a transformation of the original matrix and thus (for large problems) generating a need for huge **main memory**, or to use an **iterative** solver which works with the original matrix.

The same rationale holds for the more specialized field of **parallel linear algebra** methods. There the additional aspects originating from the parallel **computer architecture** have to be taken into account in order to choose a suitable algorithm. Several topics influencing the choice and even more the consequent implementation of these algorithms are[7,8]:

- memory architecture (**shared-memory** vs. **distributed memory**)
- amount of **memory per process/processor**
- implemented **cache** structures

It is far beyond the scope of this contribution to give an overview of the available algorithms. Instead we refer to review articles like Refs. 9–11.

From a practical point of view another important decision is whether the user implements the linear algebra algorithm **himself** or relies on **available software / libraries**. A variety of well-known, robust packages providing high computational performance are on the market, which can be used as building blocks for an application software. Some **freely-available** libraries are:

- Basic Linear Algebra Subprograms **(BLAS)**[12]
- Linear Algebra Package **(LAPACK)**[13]
- Scalable LAPACK **(ScaLAPACK)**[14]
- **(P)ARPACK** - a (parallel) package for the solution of large eigenvalue problems[15]
- Portable, Extensible Toolkit for Scientific computation **(PETSc)**[16]

Some of them like BLAS or LAPACK are **serial** software, which help to gain good single processor performance, but leave the task of **parallelization** of the high-level linear algebra computations, e.g. solution of the coupled linear equations, to the user; others, e.g. ScaLAPACK or PARPACK, contain implementations of **parallel solvers**. Thus these packages relieve the user of the parallelization, but still they rely on special data distribution schemes[17] which require a specific organization of the application program. As a consequence the user has to handle the corresponding data distribution on his own, i.e. he has to parallelize his program at least partly. Nevertheless this might be a lot easier than to implement the full parallel linear algebra algorithm.

Since both strategies are preferable under certain circumstances, we will present in the following two simple physical problems where the **parallel numerical solution** will be demonstrated paradigmatically along the two different approaches:

In Section 3 the Poisson equation will be treated using a **parallel Jacobi solver** for the evolving system of linear equations.

In Section 4 the eigenvalue problem arising from the calculation of the vibration of a membrane is solved using a **ScaLAPACK routine**.

Of course, one would not use these solutions in real applications. Neither is the Jacobi algorithm a state-of-the-art method for the solution of a system of linear equations, nor is the eigensolver from ScaLAPACK the optimal choice for the given problem. Both examples result in a sparse matrix as will be shown in the following. ScaLAPACK contains solvers for full and banded systems, whereas (P)ARACK is a library based on the Arnoldi method which is very suitable for the calculation of a **few** eigenvalues for large **sparse** systems; thus (P)ARPACK would be the natural choice for this kind of problem.

Nevertheless due to the importance of ScaLAPACK for many application fields, e.g. **multiscale simulations**, and the simplicity of the Jacobi algorithm we present them as illustrative examples.

3 The Poisson Problem

In this section we discuss the numerical solution of the Poisson equation as an example for the approximate treatment of partial differential equations. We give a short outline of

the steps necessary to obtain a serial and later on parallel implementation of the numerical solver. Similar but more elaborate material on this topic can be found in the Refs. 18–20.

In a first step we discuss the discretization of the Poisson equation and introduce one simple solver for the evolving system of linear equations. Afterwards we focus on the parallelization of this algorithm.

3.1 Discretization of the Poisson Equation

The **Poisson equation** in two dimensions is given by

$$\Delta u \;=\; \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \;=\; f(x,y) \quad , \quad (x,y) \in \Omega \subset \mathbb{R}^2 \tag{1}$$

where Ω is a domain in $\mathbb{R}^2$. For simplicity $u(x,y)$ shall be given on the boundary $\partial\Omega$ by a **Dirichlet boundary condition**

$$u(x,y) = g(x,y) \quad , \quad (x,y) \in \partial\Omega \tag{2}$$

The functions $f(x,y)$ and $g(x,y)$ are given and $u(x,y)$ is to be calculated.

Since the analytic solution of such a partial differential equation might not be feasible depending on the shape of the domain, the functions f, g etc., one often has to resort to the numerical solution of such a differential equation. In the following we will develop a simple scheme how to calculate u approximately. For this we assume that the domain has a simple form: Ω is a rectangle (Figure 1). In order to determine the approximate solution

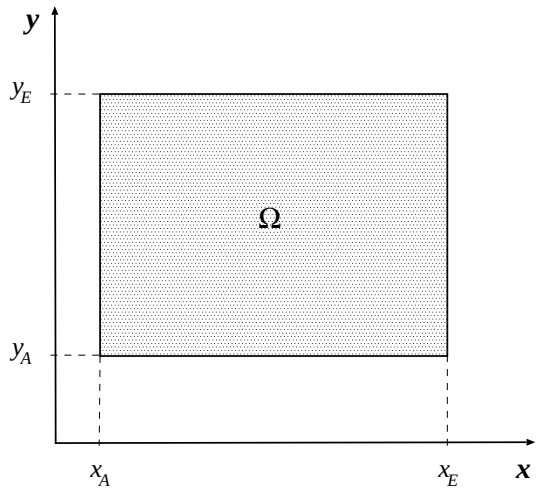


Figure 1. Rectangular domain in $\mathbb{R}^2$

of the Poisson equation, u is calculated at certain points of the rectangle. We impose $\Omega = (x_A, x_E) \times (y_A, y_E)$ with an equidistant mesh (Figure 2), where (x_A, x_E) is divided

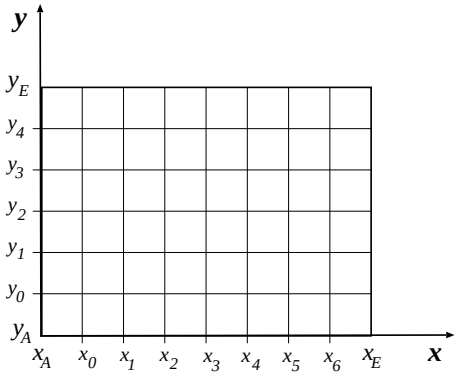


Figure 2. Mesh for $NI = 7$ and $NJ = 5$

into $(NI + 1)$ sub-intervals and (y_A, y_E) into $(NJ + 1)$ sub-intervals, $(NI, NJ \in \mathbb{N})$. The mesh width h is then given by

$$h \;=\; \frac{(x_E - x_A)}{(NI+1)} \;=\; \frac{(y_E - y_A)}{(NJ+1)} \tag{3}$$

With this choice for the mesh the approximate solution will be calculated at the $NI \cdot NJ$ inner points of the domain (The outer points don't have to be calculated, because they are given by the Dirichlet boundary condition!).

As a next step the second derivatives are replaced by finite differences. For this purpose we use the following Taylor expansions of u at a point (x, y):

$$u(x+h,y) = u(x,y) + hu_x(x,y) + \frac{h^2}{2!}u_{xx}(x,y) + \frac{h^3}{3!}u_{xxx}(x,y) \pm \dots \tag{4}$$

$$u(x-h,y) = u(x,y) - hu_x(x,y) + \frac{h^2}{2!}u_{xx}(x,y) - \frac{h^3}{3!}u_{xxx}(x,y) \pm \dots \tag{5}$$

Addition of both equations and division by h^2 gives

$$\frac{u(x-h,y) - 2u(x,y) + u(x+h,y)}{h^2} = u_{xx}(x,y) + O(h^2) \tag{6}$$

The result of the analogous procedure for the y-direction is

$$\frac{u(x,y-h) - 2u(x,y) + u(x,y+h)}{h^2} = u_{yy}(x,y) + O(h^2) \tag{7}$$

Using these finite differences the Poisson equation for the $NI \cdot NJ$ inner mesh points of the domain Ω is given by

$$\begin{gathered} u_{xx}(x_i,y_j) + u_{yy}(x_i,y_j) = f(x_i,y_j) \\ (i = 0,\dots,NI-1\;;\; j = 0,\dots,NJ-1) \end{gathered} \tag{8}$$

By neglecting the discretization error $O(h^2)$ Eqs. (8) can be written as:

$$u_{i,j-1} + u_{i-1,j} - 4u_{i,j} + u_{i,j+1} + u_{i+1,j} = h^2 f_{i,j} \tag{9}$$

for $i = 0, \ldots, NI-1\ ;\ j = 0, \ldots, NJ-1$. The unknowns

$$u_{i,j} := u(x_i, y_j) \tag{10}$$

have to be calculated from the $NI \cdot NJ$ coupled linear equations (9).

The approximation used here for $u_{xx} + u_{yy}$ is called **5-point stencil** (Figure 3). The

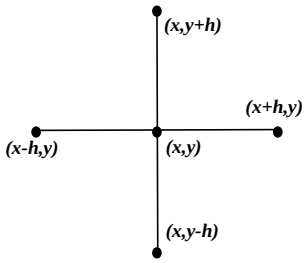


Figure 3. 5-point stencil

name describes the numerical dependency between the points of the mesh. The lexicographical numbering (Figure 4) of the mesh points

$$l = j \cdot NI + i + 1 \quad ; i = 0, \ldots, NI-1\ ;\ j = 0, \ldots, NJ-1 \tag{11}$$

and

$$u_l \ := \ u_{i,j} \tag{12}$$

allows a compact representation of the system of linear equations by means of a matrix.

The coefficient matrix A is a block tridiagonal matrix:

$$A = \begin{pmatrix} A_1 & I & & & \\ I & A_2 & I & & \\ & \ddots & \ddots & \ddots & \\ & & I & A_{NJ-1} & I \\ & & & I & A_{NJ} \end{pmatrix} \in \mathbb{R}^{(NI \cdot NJ) \times (NI \cdot NJ)} \tag{13}$$

with $A_i, I \in \mathbb{R}^{NI \times NI}$; here I is the unit matrix and

$$A_i = \begin{pmatrix} -4 & 1 & & & \\ 1 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -4 & 1 \\ & & & 1 & -4 \end{pmatrix} \quad i = 1, \ldots, NJ \tag{14}$$

This means the task to solve the Poisson equation numerically leads us to the problem to find the solution of a system of linear equations:

$$A\,\vec{u} \ = \ \vec{b} \tag{15}$$

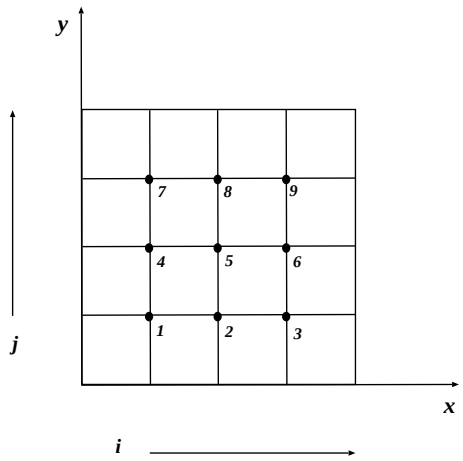


Figure 4. Lexicographical numbering of a 5×5 mesh (with 3×3 inner points)

with

$$A \in \mathbb{R}^{(NI \cdot NJ) \times (NI \cdot NJ)} \quad \text{and} \quad \vec{u}, \vec{b} \in \mathbb{R}^{(NI \cdot NJ)} \tag{16}$$

The right hand side $\vec{b}$ contains the $f_{i,j}$ of the differential equations as well as the Dirichlet boundary condition.

For the solution of these coupled linear equations many well-known numerical algorithms are available. We will focus here on the classic but very simple Jacobi algorithm.

3.2 The Jacobi Algorithm for Systems of Linear Equations

Suppose

$$A \;=\; D - L - U \tag{17}$$

is a decomposition of the matrix A, where D is the diagonal sub-matrix, $-L$ is the strict lower triangular part and $-U$ the strict upper triangular part. Then for the system of linear equations holds

$$A\,\vec{u} = \vec{b} \quad \Leftrightarrow \quad (D - L - U)\,\vec{u} = \vec{b} \quad \Leftrightarrow \quad D\,\vec{u} = (L + U)\,\vec{u} + \vec{b} \quad \Leftrightarrow \tag{18}$$

$$\vec{u} \;=\; D^{-1}(L + U)\,\vec{u} + D^{-1}\,\vec{b} \quad \text{if } D^{-1} \text{ exists .} \tag{19}$$

From Eq. (19) follows the iteration rule (for D non-singular)

$$\vec{u}^{(k)} = D^{-1}(L + U)\,\vec{u}^{(k-1)} + D^{-1}\,\vec{b} \quad \text{with } k = 1, 2, \ldots \tag{20}$$

This iterative procedure is known as **Jacobi** or **total-step method**. The second name is motivated by the fact that the next iteration is calculated **only** from the values of the unknowns of the last iteration. There are other schemes, e.g. Gauß-Seidel algorithm, which depend on old **and** the current iteration of the unknowns!
The corresponding pseudo code for the serial Jacobi algorithm is given here:

Jacobi algorithm

Choose an initial vector $\vec{u}^{\,(0)} \in \mathbb{R}^n$

For $k = 1, 2, \ldots$

For $i = 1, 2, \ldots, n$

$$u_i^{(k)} = \frac{1}{a_{ii}} \left(b_i - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij}\, u_j^{(k-1)} \right)$$

The Poisson equation (9) discretized with the **5-point stencil** results in the following iteration procedure

$$\begin{pmatrix} u_1 \\ \vdots \\ \vdots \\ \vdots \\ u_N \end{pmatrix}^{(k)} = -\frac{1}{4} \begin{pmatrix} A'_1 & -I & & & \\ -I & A'_2 & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & A'_{NJ-1} & -I \\ & & & -I & A'_{NJ} \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ \vdots \\ \vdots \\ u_N \end{pmatrix}^{(k-1)} - \frac{1}{4} \begin{pmatrix} b_1 \\ \vdots \\ \vdots \\ \vdots \\ b_N \end{pmatrix} \tag{21}$$

with $N = NI \cdot NJ$ and

$$A'_i = \begin{pmatrix} 0 & -1 & & & \\ -1 & 0 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & -1 \\ & & & -1 & 0 \end{pmatrix} \quad i = 1, \ldots, NJ \tag{22}$$

This can be seen easily by application of the Jacobi matrix decomposition on the coefficient matrix given by Eqs. (13) and (14). The pseudo code for this special case is shown here

Jacobi algorithm for the Poisson equation

Choose initial vector $\vec{u}^{\,(0)} \in \mathbb{R}^N$

For $k = 1, 2, \ldots$

For $j = 0, 1, \ldots, NJ - 1$

For $i = 0, 1, \ldots, NI - 1$

$$u_{i,j}^{(k)} = \frac{1}{4} \left(u_{i,j-1}^{(k-1)} + u_{i-1,j}^{(k-1)} + u_{i,j+1}^{(k-1)} + u_{i+1,j}^{(k-1)} - h^2 f_{i,j} \right)$$

3.3 Parallelization of the Jacobi Algorithm

The numerical treatment of the Poisson equation led us to the task to solve a system of linear equations. We introduced the Jacobi algorithm as a simple method to calculate this solution and presented the corresponding serial pseudo code. Now the next step is to discuss strategies **how to implement the Jacobi algorithm on a parallel computer**.

The important point about the Jacobi (total-step) algorithm one has to remember is that the calculation of the new iteration only depends on the values of the unknowns from the **last iteration** as can be seen for instance from Eq. (21). As a consequence the processors of a parallel computer can calculate the new iteration of the unknowns simultaneously, supposed each unknown is assigned to its own processor. This makes the parallelization of the Jacobi algorithm quite easy compared to other methods with more complicated dependencies between different iterations.

Usually the number of unknowns is much larger than the number of available processors. Thus some / many unknowns have to be assigned to one processor, i.e. for our example: the inner points of Ω (Figure 1) are distributed to the available processors. With other words the **domain** Ω is **decomposed** according to a suitable strategy.
The criteria for a "suitable" strategy are

- **load balance**, i.e. same / similar number of unknowns for each processor
- minimization of the **communication** between the processors, i.e. the dependency on unknowns stored on other processors (within one iteration step!) is reduced

For our example, the Poisson equation in two dimensions, a reasonable domain decomposition is shown in Figure 5: Each processor "owns" a domain of the same size, i.e. each

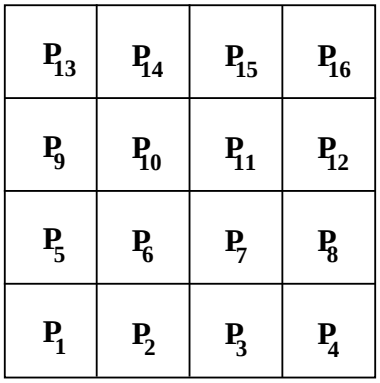


Figure 5. Domain decomposition of a square Ω with 16 processors

P_i "owns" the same number of points. Furthermore the ratio area to edges of each square and consequently the ratio between the number of inner points (no dependency on points "owned" by other processors) to the number of points near the boundary is rather good. This point can been seen even better from Figure6.

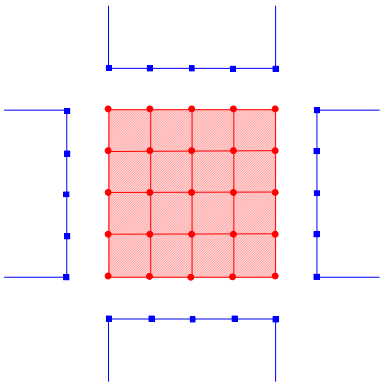


Figure 6. Dependency of the unknowns of processor P_i (red) on values stored on the neighbors (blue)

In Fig. 6 the points / corresponding unknowns of processor P_i are represented by red circles, whereas the blue squares depict the **ghost points**, i.e. points stored on other processors which are required for the calculation of the next iteration of the unknowns on processor P_i.

The dependencies / ghost points shown in Fig. 6 are a result of the 5-point stencil (see Fig. 3) originating from the Laplace operator in Eq. (1). Thus the domain decomposition of choice might differ for other differential equations or other discretization schemes, e.g. finite elements.

Due to the dependencies between the unknowns "owned" by different processors it is clear that the parallelization of the Jacobi algorithm has to introduce statements which will take care of the communication between the processors. One portable way to handle the communication is the widely used **Message Passing Interface (MPI)**[21] library.
The pseudo code of the parallel Jacobi algorithm is given here:

Parallel Jacobi algorithm

Choose initial values for the **own mesh points** *and the* **ghost points**

Choose initial Precision (*e.g. Precision* $= 10^{10}$)

While Precision $> \varepsilon$ (*e.g.* $\varepsilon = 10^{-5}$)

1. *Calculate next iteration for the* **own domain**
2. *Send the new iteration* **on boundary of domain** *to* **neighboring processors**
3. *Receive the new iteration for the* **ghost points**
4. *Calculate Precision* $= \|A\vec{u}^{(k)} - \vec{b}\|$

End While

The steps 2 and 3 show the extension of the serial Jacobi algorithm by *Send* and *Receive* statements. This is of course only one very simple way to implement such a communication with the four neighboring processors. In real applications one will look for more efficient communication patterns.

Step 4 requires implicitly **global communication**, because the vector $\vec{u}^{(k)}$ holding the approximate solution of the system of linear equations is distributed over all processors. As soon as the required precision of the solution is achieved the iteration stops.

4 Vibration of a Membrane

The vibration of a homogeneous membrane is governed by the time-dependent partial differential equation[22]

$$\frac{\partial^2 v}{\partial t^2} \;=\; \Delta\, v \tag{23}$$

In order to solve this equation we make a separation ansatz for the time and spatial variables:

$$v(x,y,t) \;=\; u(x,y)\, g(t) \tag{24}$$

By insertion of Eq. (24) into Eq. (23) one immediately obtains

$$g(t)\, \Delta\, u(x,y) \;=\; u(x,y)\, g''(t) \quad \Leftrightarrow \tag{25}$$

$$\frac{\Delta\, u(x,y)}{u(x,y)} \;=\; \frac{g''(t)}{g(t)} \tag{26}$$

The left side of Eq. (26) is independent of t, the right side of x, y. Therefore both sides must be equal to a constant $-\lambda$

$$\frac{\Delta\, u(x,y)}{u(x,y)} \;=\; \frac{g''(t)}{g(t)} \;=\; -\lambda \quad \Leftrightarrow \tag{27}$$

$$\Delta\, u(x,y) \;=\; -\lambda\, u(x,y) \quad \text{and} \quad g''(t) \;=\; -\lambda\, g(t) \tag{28}$$

The differential equation for $g(t)$ can be solved easily with the usual ansatz (a linear combination of trigonometric functions).

In the following we want to solve the spatial partial differential equation

$$\Delta\, u(x,y) \;=\; -\lambda\, u(x,y) \tag{29}$$

numerically. In section 3.1 we presented the discretization of the Poisson equation in two dimensions. In order to allow a re-use of the results derived there, we will calculate the solution of Eq. (29) for a **rectangular** membrane / domain.

Furthermore we choose for simplicity the Dirichlet boundary condition

$$u(x,y) \;=\; 0 \quad \text{for } (x,y) \in \partial\Omega \tag{30}$$

Using the same discretization for the Laplace operator and lexicographical numbering of the mesh points / unknowns as in section 3.1 one can see easily that Eq. (29) leads to the **eigenvalue problem**

$$A\, \vec{u} \;=\; -\lambda\, \vec{u} \tag{31}$$

where the matrix A is given by Eqs. (13) and (14).

In section 3 we presented a simple algorithm for the solution of the system of linear equations and discussed the parallelization by hand. For the eigenvalue problem we choose a different strategy: We make use of a widely used parallel library, namely the **ScaLAPACK** library.

4.1 Parallel Solution Using the ScaLAPACK Library

The largest and most flexible public domain library with linear algebra routines for distributed memory parallel systems up to now is ScaLAPACK[14]. Within the ScaLAPACK project many LAPACK routines were ported to distributed memory computers using MPI.

The basic routines of ScaLAPACK are the **PBLAS** (Parallel Basic Linear Algebra Subroutines)[23]. They contain parallel versions of the BLAS which are parallelized using **BLACS** (Basic Linear Algebra Communication Subprograms)[24] for communication and sequential BLAS for computation. Thus the PBLAS deliver very good performance on most parallel computers.

ScaLAPACK contains direct parallel solvers for dense linear systems (LU and Cholesky decomposition), linear systems with band matrices as well as parallel routines for the solution of linear least squares problems and for singular value decomposition.

Furthermore there are several different routines for the solution of the full symmetric eigenproblem. We will focus in the following on a **simple driver routine** using the QR-algorithm, which computes all eigenvalues and optionally all eigenvectors of the matrix.

Besides this there are other eigensolvers available which are implementations of other algorithms, e.g. a divide-and-conquer routine; an additional expert driver allows to choose a range of eigenvalues and optionally eigenvectors to be computed.

For performance and load balancing reasons ScaLAPACK uses a **two-dimensional block cyclic distribution** for full matrices (see ScaLAPACK Users' Guide)[17]:
First the matrix is divided into blocks of size MB $\times$ NB, where MB and NB are the number of rows and columns per block, respectively. These blocks are then uniformly distributed across the MP $\times$ NP **rectangular processor grid** in a cyclic manner. As a result, each process owns a collection of blocks. Figure 7 shows the distribution of a (9×9) matrix subdivided into blocks of size (3×2) distributed across a (2×2) processor grid.

	0		1		0		1		0
0	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}	a_{18}	a_{19}
	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	a_{26}	a_{27}	a_{28}	a_{29}
	a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	a_{36}	a_{37}	a_{38}	a_{39}
1	a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	a_{46}	a_{47}	a_{48}	a_{49}
	a_{51}	a_{52}	a_{53}	a_{54}	a_{55}	a_{56}	a_{57}	a_{58}	a_{59}
	a_{61}	a_{62}	a_{63}	a_{64}	a_{65}	a_{66}	a_{67}	a_{68}	a_{69}
0	a_{71}	a_{72}	a_{73}	a_{74}	a_{75}	a_{76}	a_{77}	a_{78}	a_{79}
	a_{81}	a_{82}	a_{83}	a_{84}	a_{85}	a_{86}	a_{87}	a_{88}	a_{89}
	a_{91}	a_{92}	a_{93}	a_{94}	a_{95}	a_{96}	a_{97}	a_{98}	a_{99}

Figure 7. Block cyclic 2D distribution of a (9×9) matrix subdivided into (3×2) blocks on a (2×2) processor grid. The numbers outside the matrix indicate processor row and column indices, respectively.

ScaLAPACK as a parallel successor of LAPACK attempts to leave the calling sequence of the subroutines unchanged as much as possible in comparison to the corresponding sequential subroutine from LAPACK. The user should have to change only a few parameters in the calling sequence to use ScaLAPACK routines instead of LAPACK routines.

Therefore ScaLAPACK uses so-called **descriptors**, which are integer arrays containing all necessary information about the **distribution of the matrix**. This descriptor appears in the calling sequence of the parallel routine instead of the leading dimension of the matrix in the sequential one.

For example the sequential simple driver **DSYEV** from LAPACK for the computation of **all eigenvalues** and (optionally) eigenvectors of a **real symmetric** ($N \times N$) matrix A has the following calling sequence[25]:

```
...
CALL DSYEV(JOBZ, UPLO, N, A, LDA, W, WORK, LWORK, INFO)
...
```

where JOBZ and UPLO are characters indicating whether to compute eigenvectors, and whether the lower or the upper triangular part of the matrix A is provided. LDA is the leading dimension of A and W is the array of eigenvalues of A. The other variables are used as workspace and for error report.

The corresponding ScaLAPACK routine **PDSYEV** is called as follows[17]:

```
...
      CALL PDSYEV ( JOBZ, UPLO, N, A, IA, JA, DESCA,
     $              W, Z, IZ, JZ, DESCZ, WORK, LWORK, INFO )
...
```

As one can see the leading dimension LDA of the LAPACK call is substituted by the indices IA and JA and the descriptor DESCA. IA and JA indicate the start position of the **global matrix** (usually IA, JA = 1, but in cases where the global matrix is a sub-matrix of a larger matrix IA, JA $\neq$ 1 might occur) , whereas DESCA contains all information regarding the distribution of the global matrix. The parameters IZ, JZ, and DESCZ provide the same information for Z, the matrix of the eigenvectors calculated by PDSYEV.

In order to use the ScaLAPACK routine the user has to distribute his system matrix in the way required by ScaLAPACK. Thus the user has to setup the **processor grid** by initializing MP, the number of processor rows, and NP, the number of processor columns. Furthermore one has to choose a suitable blocking of the matrix, i.e. MB and NB. For many routines, especially for the eigenvalue solvers and the Cholesky decomposition, MB=NB is mandatory. (Since MB and NB are crucial for the performance of the solver, one has to use these parameters with care.[26]) Further details on the two-dimensional block cyclic distribution of the matrix A given by Eqs. (13) and (14) can be found in the appendix.

Once the matrix has been distributed to the processors, the calculation of the eigenvalues and corresponding eigenvectors for the vibration of the rectangular membrane (Eq. (31)) can be calculated easily by one call of the routine **PDSYEV**. Please note that the matrix of the eigenvectors Z, is distributed to the processors; thus if necessary, e.g. for output, it is again the task of the user to collect the different local data and to generate the global matrix.

5 Performance of Parallel Linear Algebra Libraries

The performance, i.e. scalability, of parallel libraries and the availability of optimized implementations for specific hardware platforms are major criteria for the selection of (linear algebra) functions to be used within a scientific application. Especially the availability of optimized software differs largely for different architectures: for distributed memory systems using MPI many (non-)numerical libraries are freely-available or are provided by the vendors. For new and non-standard hardware like the **PowerXCell 8i processor**[27] the situation is not that comfortable. This can reduce the usability largely particularly for the Cell processors which have to be programmed in **assembler code style**.

At present a **BLAS** library, which can be used just like its serial counterpart and hides the complexity of the parallel hardware from the application developer[28], exists for the PowerXCell architecture whereas other linear algebra libraries are still under development. Figure 8 shows the processing power of a PowerXCell 8i processor in Gflops (= 10^9 Floating point operations per second) for a **double precision** matrix-matrix multiplication (BLAS routine **DGEMM**) as function of the matrix size. The results are shown for calculations where 2, 4 and 8 Synergistic Processing Elements **(SPE)**[27] are used; the Cell processor contains 8 SPEs which have a theoretical compute power of 12.8 Gflops each (3.2 GHz, 4 flops per cycle).

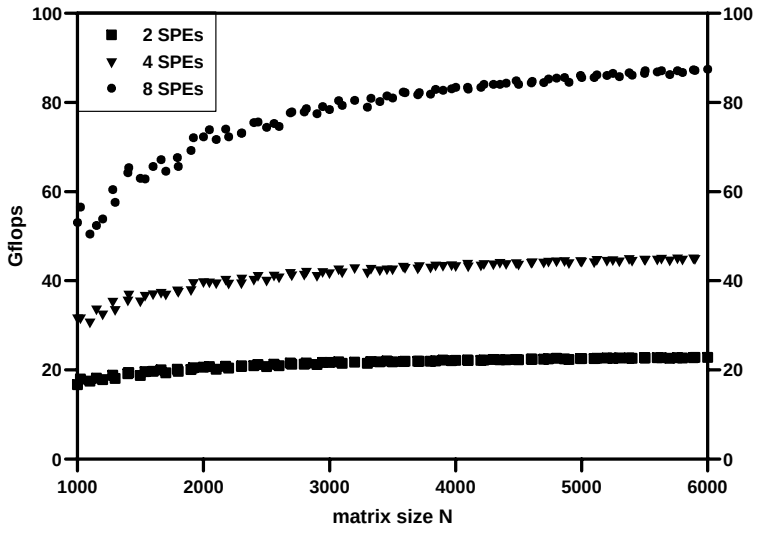


Figure 8. Processing power [Gflops] of a double precision matrix-matrix multiplication as function of the matrix size for a **PowerXCell processor**. Results are shown for different numbers of SPEs: 2, 4 and 8 SPEs.

From Figure 8 one learns that the sustained performance of a PowerXCell processor for the double precision matrix-matrix multiplication is approximately 90 Gflops and thus more than 85% of the theoretical peak performance (102.4 Gflops). Furthermore one sees that the number of computations per second scales well with the number of SPEs and is quite independent of the problem size. The recent multi-core processor **Intel Core i7**

(quad-core) has a theoretical compute power of 51.2 Gflops; the PowerXCell shows twice the performance in a real computation but for a substantially lower price and electrical power consumption.

A complementary strategy to build supercomputers is to add huge numbers of low-cost, regarding price as well as power consumption, processors to gain high compute power. This concept has been implemented for instance with IBM's Blue Gene/P where each processor core supplies 'only' 3.4 Gflops (850 MHz, 4 flops per cycle) and has a local memory of 512 MByte. But the complete Blue Gene/P system **JUGENE**[29] at JSC with 65536 cores accomplishes more the 220.000 Gflops and has a main memory of 128 TByte (= 131.072 GBytes).

Figure 9 gives the performance of JUGENE for the parallel double precision matrix-matrix multiplication **PDGEMM** from the ScaLAPACK library using 256, 1024 and 4096 of its processors. Obviously the real compute power as well as its scaling with increasing

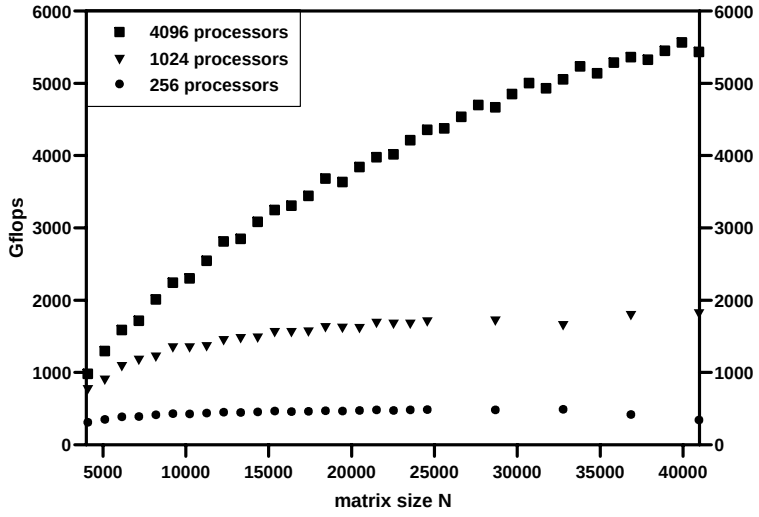


Figure 9. Processing power [Gflops] of a double precision matrix-matrix multiplication as function of the matrix size on JSC's **Blue Gene/P** system. Results are shown for different processor numbers: 256, 1024 and 4096 processors.

processor number depends largely on the problem size. This is a general observation on massively parallel computer architectures: Most algorithms show a much better parallel performance if the problem sizes increase / scale together with the number of processors employed - a behaviour known as **weak scaling** and foreseen by Gustafson[30] in 1988.

A computation rate of approximately 5.500 Gflops is shown in Figure 9 for a matrix size of 40.000 and 4096 processors. This result is only about 40% of the theoretical peak performance of the 4096 processors and the parallel speedup for increasing processor numbers is far from optimal, nevertheless it illustrates the potential of supercomputers like the IBM Blue Gene/P with several ten thousand to hundreds of thousands of processors for scientific applications. With these results in mind it is no surprise that the combination of

a massively parallel system and special high performance processors allowed to enter the era of **Petaflop computing** with the **IBM Roadrunner**[3].

6 Conclusion

In this contribution we presented two examples for the numerical treatment of partial differential equations using parallel algorithms / computers. Both problems were tackled with different strategies: The first example has been solved by means of the simple Jacobi algorithm and a suitable parallelization scheme was discussed. In the second case the parallel calculation has been performed with the help of the ScaLAPACK library.

The pros and cons of the different strategies are obvious. If a suitable parallel library is available and a reorganization of the application software according to the complex data distribution schemes of the libraries is possible, the parallel routines from the library will provide a robust numerical solution with fair or even good performance. Otherwise the user has to choose a parallelization scheme which best fits his specific application problem and he has to implement the necessary algorithms himself; in order to improve the single processor performance it is still recommendable to use serial library routines, e.g. from BLAS or LAPACK, wherever possible!

Benchmarks of a parallel linear algebra routine were shown for the IBM Blue Gene/P architecture and for the PowerXCell processor. The results demonstrate the compute power of special purpose processors as well as the potential of massively parallel computers.

Appendix

In section 4.1 some information on the **two-dimensional block cyclic distribution** of the data used by **ScaLAPACK** has been given. In this appendix we will discuss this issue in greater detail.

In Fig. 10 a code fragment is shown which distributes the N $\times$ N matrix given by Eqs. (13) and (14) according to the ScaLAPACK data scheme with block sizes NB=MB to an MP $\times$ NP processor grid. Inclusion of this fragment into a parallel program allows the calculation of the eigenvalues and eigenvectors using the routine **PDSYEV**:

```
...
      CALL PDSYEV ( JOBZ, UPLO, N, A, IA, JA, DESCA,
     $              W, Z, IZ, JZ, DESCZ, WORK, LWORK, INFO )
...
```

Notice that in the sequential as well as in the parallel routine the matrix A is destroyed. The difference is that in the sequential case if the eigenvectors are requested A is overwritten by the eigenvectors whereas in the parallel case the eigenvectors are stored to a separate matrix Z.

The matrix Z has to be allocated with the same local sizes as A and DESCZ is filled with the same values as DESCA. The size LWORK of the local workspace WORK can be found in the ScaLAPACK Users' Guide.

```
1  ! Create the MP * NP processor grid
2  CALL BLACS_GRIDINIT(ICTXT,'Row-major',MP,NP)
3  ! Find my processor coordinates MYROW and MYCOL
4  ! NPROW returns the same value as MP,
5  ! NPCOL returns the same value as NP
6  CALL BLACS_GRIDINFO(ICTXT,NPROW,NPCOL,MYROW,MYCOL)
7  ! Compute local dimensions with routine NUMROC from TOOLS
8  ! N is dimension of the matrix
9  ! NB is block size
10 MYNUMROWS = NUMROC(N,NB,MYROW,0,NPROW)
11 MYNUMCOLS = NUMROC(N,NB,MYCOL,0,NPCOL)
12 ! Local leading dimension of A,
13 ! Number of local rows of A
14 MXLLDA = MYNUMROWS
15 ! Allocate only the local part of A
16 ALLOCATE(A(MXLLDA,MYNUMCOLS))
17 ! Fill the descriptors, P0 and Q0 are processor coordinates
18 ! of the processor holding global element A(1,1)
19 CALL DESCINIT(DESCA,N,N,NB,NB,P0,Q0,ICTXT,MXLLDA,INFO)
20 ! Fill the local part of the matrix with data
21 do j = 1, MYNUMCOLS, NB        ! Fill the local column blocks
22   do jj=1,min(NB,MYNUMCOLS-j+1)   ! all columns of one block
23     jloc = j-1 + jj
   ! local column index
24     jglob = (j-1)*NPCOL + MYCOL*NB +jj ! global column index
25     do i = 1, MYNUMROWS, NB      ! local row blocks in column
26       do ii=1,min(NB,MYNUMROWS-i+1)
   ! rows in row block
27         iloc = i-1 + ii
   ! local row index
28         iglob = (i-1)*NPROW + MYROW*NB+ii ! global row index
29         A(iloc,jloc) = 0
30         If (iglob==jglob) A(iloc,jloc)=-4
31         If (iglob==jglob+1.and.mod(jglob,NI)/=0) &
32                                              A(iloc,jloc)=1
33         If (jglob==iglob+1.and.mod(iglob,NI)/=0) &
34                                              A(iloc,jloc)=1
35         If (iglob==jglob+NI) A(iloc,jloc)=1
36         If (jglob==iglob+NI) A(iloc,jloc)=1
37       enddo
38     enddo
39   enddo
40 enddo
```

Figure 10. Code fragment which distributes the matrix given by Eqs. (13) and (14) according to ScaLAPACK (It is assumed that MB=NB=NI and N=NI·NJ).

The four nested loops in Fig. 10 show how local and global indices can be computed from block sizes, the number of rows and columns in the processor grid and the processor coordinates. The conversion of global to local indices and vice versa is supported by some auxiliary routines in the **TOOLS** sub-library of ScaLAPACK. Here the routine NUMROC is used to calculate the number of rows / columns stored on the corresponding processor.

There is also a sub-library **REDIST** of ScaLAPACK which allows the redistribution of any two-dimensional block cyclically distributed matrix to any other block cyclic two-dimensional distribution. Thus if A was column cyclically distributed or if the eigenvectors have to be column cyclically distributed for further computations they can be redistributed by such a routine, as a column cyclic distribution is nothing else but a block cyclic two-dimensional distribution to a $1 \times$ NPR (with NPR = number of processors) grid with block size 1.

References

1. IBM System Blue Gene/P
   `http://www-03.ibm.com/systems/deepcomputing/bluegene/`
2. IBM Cell Broadband Engine technology
   `http://www-03.ibm.com/technology/cell/`
3. Los Alamos National Lab's Supercomputer Roadrunner
   `http://www.lanl.gov/roadrunner/`
4. Scalable Performance Analysis of Large-Scale Applications (Scalasca)
   `http://www.fz-juelich.de/jsc/scalasca/`
5. J. J. Dongarra, I. S. Duff, D. .C. Sorensen, and H. A. van der Vorst, *Numerical Linear Algebra for High-Performance Computers*, SIAM, Philadelphia (1998).
6. `http://gams.nist.gov/Classes.html` and especially
   `http://gams.nist.gov/serve.cgi/Class/D/`
7. B. Wilkinson and M. Allen, *Parallel Programming: Techniques and Applications using networked Workstations and Parallel Computers*, Pearson, Upper Saddle River (2005).
8. A. Grama, A. Gupta, G. Karypis, and V. Kumar, *Introduction to Parallel Computing*, Pearson, Harlow (2003).
9. M. Bücker, *Iteratively Solving Large Sparse Linear Systems on Parallel Computers* in Quantum Simulation of Complex Many-Body Systems: From Theory to Algorithms, J. Grotendorst et al. (Ed.), John von Neumann Institute for Computing, NIC Series Vol. **10**, 521-548 (2002)
   `http://www.fz-juelich.de/nic-series/volume10/buecker.pdf`
10. B. Lang, *Direct Solvers for Symmetric Eigenvalue Problems* in Modern Methods and Algorithms of Quantum Chemistry, J. Grotendorst (Ed.), John von Neumann Institute for Computing, NIC Series Vol. **3**, 231-259 (2000).
    `http://www.fz-juelich.de/nic-series/Volume3/lang.pdf`
11. B. Steffen, *Subspace Methods for Sparse Eigenvalue Problems* in Modern Methods and Algorithms of Quantum Chemistry, J. Grotendorst (Ed.), John von Neumann Institute for Computing, NIC Series Vol. **3**, 307-314 (2000).
    `http://www.fz-juelich.de/nic-series/Volume3/steffen.pdf`

12. Basic Linear Algebra Subprograms (BLAS)
http://www.netlib.org/blas/
13. Linear Algebra Package (LAPACK)
http://www.netlib.org/lapack/
14. Scalable Linear Algebra Package (ScaLAPACK)
http://www.netlib.org/scalapack/
15. http://www.caam.rice.edu/software/ARPACK/
16. Portable, Extensible Toolkit for Scientific computation (PETSc)
http://www-unix.mcs.anl.gov/petsc/petsc-as/
17. L. S. Blackford, J. Choi, A. Cleary et al., *ScaLAPACK Users' Guide*, SIAM, Philadelphia (1997).
18. G. Ahlefeld, I. Lenhardt, and H. Obermaier, *Parallele numerische Algorithmen*, Springer, Berlin (2002).
19. J. M. Ortega, *Introduction to parallel and vector solution of linear systems*, Plenum Press, New York (1988).
20. J. Zhu, *Solving partial differential equations on parallel computers*, World Scientific, Singapore (1994).
21. W. Gropp, E. Lus,k and A. Skjellum, *Using MPI : Portable Parallel Programming with the Message-Passing Interface*, MIT Press, Cambridge (1994).
22. R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Volume I, Interscience Publishers, New York (1953).
23. Parallel Basic Linear Algebra Subprograms (PBLAS)
http://www.netlib.org/scalapack/pblas_qref.html
24. J. J. Dongarra and R. C. Whaley, *A User's Guide to the BLACS*, LAPACK Working Note **94** (1997).
http://www.netlib.org/lapack/lawns/lawn94.ps
25. E. Anderson, Z. Bai, C. Bischof et al., *LAPACK Users' Guide, Second Edition*, SIAM, Philadelphia (1995).
26. I. Gutheil, *Basic Numerical Libraries for Parallel Systems* in Modern Methods and Algorithms of Quantum Chemistry, J. Grotendorst (Ed.), John von Neumann Institute for Computing, NIC Series Vol. **3**, 47-65 (2000).
http://www.fz-juelich.de/nic-series/Volume3/gutheil.pdf
27. IBM PowerXCell 8i processor datasheet
http://www-03.ibm.com/technology/resources/
28. IBM Cell Broadband Engine - Software Development Kit (SDK) 3.1
http://www-128.ibm.com/developerworks/power/cell/index.html
29. Blue Gene/P at JSC
http://www.fz-juelich.de/jsc/jugene/
30. John L. Gustafson, *Reevaluating Amdahl's law*, Communications of the ACM **31**, 532–533, 1988.

Already published:

Modern Methods and Algorithms of Quantum Chemistry - Proceedings
Johannes Grotendorst (Editor)
Winter School, 21 - 25 February 2000, Forschungszentrum Jülich
NIC Series Volume 1
ISBN 3-00-005618-1, February 2000, 562 pages
out of print

Modern Methods and Algorithms of Quantum Chemistry - Poster Presentations
Johannes Grotendorst (Editor)
Winter School, 21 - 25 February 2000, Forschungszentrum Jülich
NIC Series Volume 2
ISBN 3-00-005746-3, February 2000, 77 pages
out of print

Modern Methods and Algorithms of Quantum Chemistry - Proceedings, Second Edition
Johannes Grotendorst (Editor)
Winter School, 21 - 25 February 2000, Forschungszentrum Jülich
NIC Series Volume 3
ISBN 3-00-005834-6, December 2000, 638 pages
out of print

Nichtlineare Analyse raum-zeitlicher Aspekte der hirnelektrischen Aktivität von Epilepsiepatienten
Jochen Arnold
NIC Series Volume 4
ISBN 3-00-006221-1, September 2000, 120 pages

Elektron-Elektron-Wechselwirkung in Halbleitern: Von hochkorrelierten kohärenten Anfangszuständen zu inkohärentem Transport
Reinhold Lövenich
NIC Series Volume 5
ISBN 3-00-006329-3, August 2000, 146 pages

Erkennung von Nichtlinearitäten und wechselseitigen Abhängigkeiten in Zeitreihen
Andreas Schmitz
NIC Series Volume 6
ISBN 3-00-007871-1, May 2001, 142 pages

Multiparadigm Programming with Object-Oriented Languages - Proceedings
Kei Davis, Yannis Smaragdakis, Jörg Striegnitz (Editors)
Workshop MPOOL, 18 May 2001, Budapest
NIC Series Volume 7
ISBN 3-00-007968-8, June 2001, 160 pages

Europhysics Conference on Computational Physics - Book of Abstracts
Friedel Hossfeld, Kurt Binder (Editors)
Conference, 5 - 8 September 2001, Aachen
NIC Series Volume 8
ISBN 3-00-008236-0, September 2001, 500 pages

NIC Symposium 2001 - Proceedings
Horst Rollnik, Dietrich Wolf (Editors)
Symposium, 5 - 6 December 2001, Forschungszentrum Jülich
NIC Series Volume 9
ISBN 3-00-009055-X, May 2002, 514 pages

Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms - Lecture Notes
Johannes Grotendorst, Dominik Marx, Alejandro Muramatsu (Editors)
Winter School, 25 February - 1 March 2002, Rolduc Conference Centre, Kerkrade, The Netherlands
NIC Series Volume 10
ISBN 3-00-009057-6, February 2002, 548 pages

Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms- Poster Presentations
Johannes Grotendorst, Dominik Marx, Alejandro Muramatsu (Editors)
Winter School, 25 February - 1 March 2002, Rolduc Conference Centre, Kerkrade, The Netherlands
NIC Series Volume 11
ISBN 3-00-009058-4, February 2002, 194 pages

Strongly Disordered Quantum Spin Systems in Low Dimensions: Numerical Study of Spin Chains, Spin Ladders and Two-Dimensional Systems
Yu-cheng Lin
NIC Series Volume 12
ISBN 3-00-009056-8, May 2002, 146 pages

Multiparadigm Programming with Object-Oriented Languages - Proceedings
Jörg Striegnitz, Kei Davis, Yannis Smaragdakis (Editors)
Workshop MPOOL 2002, 11 June 2002, Malaga
NIC Series Volume 13
ISBN 3-00-009099-1, June 2002, 132 pages

Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms - Audio-Visual Lecture Notes
Johannes Grotendorst, Dominik Marx, Alejandro Muramatsu (Editors)
Winter School, 25 February - 1 March 2002, Rolduc Conference Centre, Kerkrade, The Netherlands
NIC Series Volume 14
ISBN 3-00-010000-8, November 2002, DVD

Numerical Methods for Limit and Shakedown Analysis
Manfred Staat, Michael Heitzer (Eds.)
NIC Series Volume 15
ISBN 3-00-010001-6, February 2003, 306 pages

Design and Evaluation of a Bandwidth Broker that Provides Network Quality of Service for Grid Applications
Volker Sander
NIC Series Volume 16
ISBN 3-00-010002-4, February 2003, 208 pages

Automatic Performance Analysis on Parallel Computers with SMP Nodes
Felix Wolf
NIC Series Volume 17
ISBN 3-00-010003-2, February 2003, 168 pages

Haptisches Rendern zum Einpassen von hochaufgelösten Molekülstrukturdaten in niedrigaufgelöste Elektronenmikroskopie-Dichteverteilungen
Stefan Birmanns
NIC Series Volume 18
ISBN 3-00-010004-0, September 2003, 178 pages

Auswirkungen der Virtualisierung auf den IT-Betrieb
Wolfgang Gürich (Editor)
GI Conference, 4 - 5 November 2003, Forschungszentrum Jülich
NIC Series Volume 19
ISBN 3-00-009100-9, October 2003, 126 pages

NIC Symposium 2004
Dietrich Wolf, Gernot Münster, Manfred Kremer (Editors)
Symposium, 17 - 18 February 2004, Forschungszentrum Jülich
NIC Series Volume 20
ISBN 3-00-012372-5, February 2004, 482 pages

Measuring Synchronization in Model Systems and Electroencephalographic Time Series from Epilepsy Patients
Thomas Kreutz
NIC Series Volume 21
ISBN 3-00-012373-3, February 2004, 138 pages

Computational Soft Matter: From Synthetic Polymers to Proteins - Poster Abstracts
Norbert Attig, Kurt Binder, Helmut Grubmüller, Kurt Kremer (Editors)
Winter School, 29 February - 6 March 2004, Gustav-Stresemann-Institut Bonn
NIC Series Volume 22
ISBN 3-00-012374-1, February 2004, 120 pages

Computational Soft Matter: From Synthetic Polymers to Proteins - Lecture Notes
Norbert Attig, Kurt Binder, Helmut Grubmüller, Kurt Kremer (Editors)
Winter School, 29 February - 6 March 2004, Gustav-Stresemann-Institut Bonn
NIC Series Volume 23
ISBN 3-00-012641-4, February 2004, 440 pages

Synchronization and Interdependence Measures and their Applications to the Electroencephalogram of Epilepsy Patients and Clustering of Data
Alexander Kraskov
NIC Series Volume 24
ISBN 3-00-013619-3, May 2004, 106 pages

High Performance Computing in Chemistry
Johannes Grotendorst (Editor)
Report of the Joint Research Project:
High Performance Computing in Chemistry - HPC-Chem
NIC Series Volume 25
ISBN 3-00-013618-5, December 2004, 160 pages

Zerlegung von Signalen in unabhängige Komponenten: Ein informationstheoretischer Zugang
Harald Stögbauer
NIC Series Volume 26
ISBN 3-00-013620-7, April 2005, 110 pages

Multiparadigm Programming 2003
Joint Proceedings of the
3rd International Workshop on Multiparadigm Programming with Object-Oriented Languages (MPOOL'03)
and the
1st International Workshop on Declarative Programming in the Context of Object-Oriented Languages (PD-COOL'03)
Jörg Striegnitz, Kei Davis (Editors)
NIC Series Volume 27
ISBN 3-00-016005-1, July 2005, 300 pages

Integration von Programmiersprachen durch strukturelle Typanalyse und partielle Auswertung
Jörg Striegnitz
NIC Series Volume 28
ISBN 3-00-016006-X, May 2005, 306 pages

OpenMolGRID - Open Computing Grid for Molecular Science and Engineering
Final Report
Mathilde Romberg (Editor)
NIC Series Volume 29
ISBN 3-00-016007-8, July 2005, 86 pages

GALA Grünenthal Applied Life Science Analysis
Achim Kless and Johannes Grotendorst (Editors)
NIC Series Volume 30
ISBN 3-00-017349-8, November 2006, 204 pages

Computational Nanoscience: Do It Yourself! Lecture Notes
Johannes Grotendorst, Stefan Blügel, Dominik Marx (Editors)
Winter School, 14. - 22 February 2006, Forschungszentrum Jülich
NIC Series Volume 31
ISBN 3-00-017350-1, February 2006, 528 pages

NIC Symposium 2006 - Proceedings
G. Münster, D. Wolf, M. Kremer (Editors)
Symposium, 1 - 2 March 2006, Forschungszentrum Jülich
NIC Series Volume 32
ISBN 3-00-017351-X, February 2006, 384 pages

Parallel Computing: Current & Future Issues of High-End Computing
Proceedings of the International Conference ParCo 2005
G.R. Joubert, W.E. Nagel, F.J. Peters,
O. Plata, P. Tirado, E. Zapata (Editors)
NIC Series Volume 33
ISBN 3-00-017352-8, October 2006, 930 pages

From Computational Biophysics to Systems Biology 2006 Proceedings
U.H.E. Hansmann, J. Meinke, S. Mohanty, O. Zimmermann (Editors)
NIC Series Volume 34
ISBN-10 3-9810843-0-6, ISBN-13 978-3-9810843-0-6,
September 2006, 224 pages

Dreistufig parallele Software zur Parameteroptimierung von Support-Vektor-Maschinen mit kostensensitiven Gütemaßen
Tatjana Eitrich
NIC Series Volume 35
ISBN 978-3-9810843-1-3, March 2007, 262 pages

From Computational Biophysics to Systems Biology (CBSB07) Proceedings
U.H.E. Hansmann, J. Meinke, S. Mohanty, O. Zimmermann (Editors)
NIC Series Volume 36
ISBN 978-3-9810843-2-0, August 2007, 330 pages

Parallel Computing: Architectures, Algorithms and Applications - Book of Abstracts
Book of Abstracts, ParCo 2007 Conference, 4. - 7. September 2007
G.R. Joubert, C. Bischof, F. Peters, T. Lippert, M. Bücker, P. Gibbon, B. Mohr (Eds.) NIC Series Volume 37
ISBN 978-3-9810843-3-7, August 2007, 216 pages

Parallel Computing: Architectures, Algorithms and Applications - Proceedings
Proceedings, ParCo 2007 Conference, 4. - 7. September 2007
C. Bischof, M. Bücker, P. Gibbon, G.R. Joubert, T. Lippert, B. Mohr, F. Peters (Eds.) NIC Series Volume 38
ISBN 978-3-9810843-4-4, December 2007, 830 pages

NIC Symposium 2008 - Proceedings
G. Münster, D. Wolf, M. Kremer (Editors)
Symposium, 20 - 21 February 2008, Forschungszentrum Jülich
NIC Series Volume 39
ISBN 978-3-9810843-5-1, February 2008, 380 pages

From Computational Biophysics to Systems Biology (CBSB08)
Proceedings
Ulrich H.E. Hansmann, Jan H. Meinke, Sandipan Mohanty, Walter Nadler, Olav Zimmermann (Eds.) (Editors)
NIC Series Volume 40
ISBN 978-3-9810843-6-8, July 2008, 452 pages

Multigrid Methods for Structured Grids and Their Application in Particle Simulation
Matthias Bolten
NIC Series Volume 41
ISBN 978-3-9810843-7-5, February 2009, 140 pages

All volumes are available online at

http://www.fz-juelich.de/nic-series/.

The Institute for Advanced Simulation (IAS) combines the Jülich simulation sciences and the supercomputer facility in one organizational unit. It includes those parts of the scientific institutes at Forschungszentrum Jülich which use simulation on supercomputers as their main research methodology. The Jülich Supercomputing Centre (JSC) has been an integral part of IAS since January 2008.

The Jülich Supercomputing Centre provides supercomputer resources, IT tools, methods and know-how for the Forschungszentrum Jülich, the Helmholtz Association, and – through the John von Neumann Institute for Computing – for computational scientists at German and European universities, research institutions and in industry. It operates the supercomputers, central server systems, and communication infrastructure in Jülich.

The John von Neumann Institute for Computing (NIC) is a joint foundation of Forschungszentrum Jülich, Deutsches Elektronen-Synchrotron DESY and GSI Helmholtzzentrum für Schwerionenforschung to support the supercomputer-oriented simulation sciences. The core task of NIC is the peer-reviewed allocation of supercomputing resources to computational science projects in Germany and Europe. The NIC partners also support research groups, dedicated to supercomputer-based investigation in selected fields of physics and natural sciences.

NIC Series Volume 42
ISBN 978-3-9810843-8-2

JÜLICH
FORSCHUNGSZENTRUM