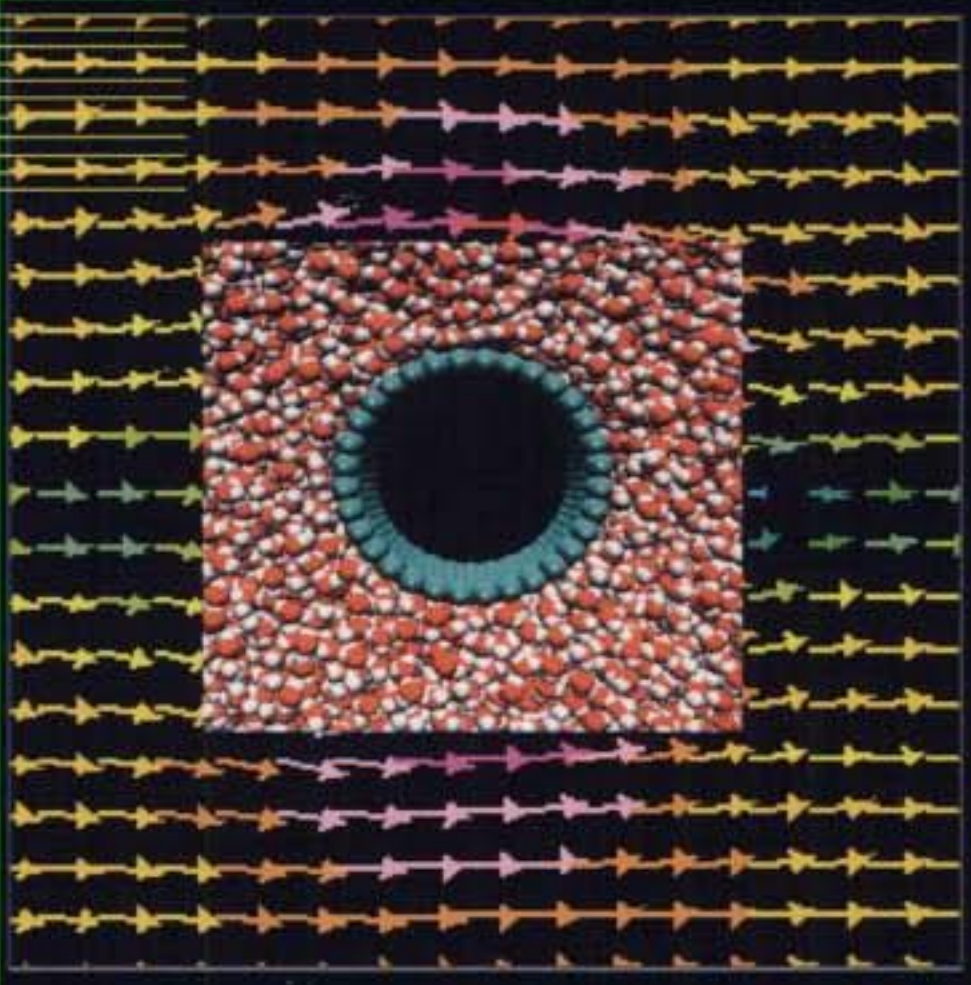
Lecture Notes in Computational Science and Engineering

39



Sabine Attinger
Petros Koumoutsakos
Editors

Multiscale Modelling and Simulation

Springer

Lecture Notes
in Computational Science
and Engineering 39

Sabine Attinger
Petros Koumoutsakos
Editors

Multiscale Modelling and Simulation

Springer

Editors

Sabine Attinger

ETHZ Computational Laboratory (CoLab)
Hirschengraben 84
Swiss Federal Institute of Technology
8092 Zürich, Switzerland
Now at:
Computational Environmental Systems
UFZ Centre for Environmental Research
Permoserstraße 15
04318 Leipzig, Germany
e-mail: sabine.attinger@ufz.de

Petros Koumoutsakos

ETHZ Computational Laboratory (CoLab)
Hirschengraben 84
Swiss Federal Institute of Technology
8092 Zürich, Switzerland
e-mail: petros@inf.ethz.ch

Library of Congress Control Number: 2004104813

Mathematics Subject Classification (2000): 65-06, 65-02, 65Z05, 65M99, 68U20, 60-08

ISSN 1439-7358
ISBN 3-540-21180-2 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in Germany

Cover Design: Friedhelm Steinen-Broo, Estudio Calamar, Spain
Cover production: *design & production*
Typeset by the authors using a Springer TEX macro package

Printed on acid-free paper 46/3142/LK - 5 4 3 2 1 0

Preface

These Lecture Series in Computational Science and Engineering include the proceedings of a Summer Program in Multi-Scale Modelling and Simulation held in Lugano, Switzerland, between August 4th and 30th, 2003. The workshop was organized by the ETHZ Computational Laboratory (CoLab) together with the Swiss Center for Scientific Computing in Manno (CSCS) and the Universit della Svizzera Italiana (USI) in the context of the ETHZ initiative in Computational Science and Engineering.

Computational Science and Engineering entails interdisciplinary research, by a sustained interaction of a critical mass of researchers, tackling complex problems of scientific and engineering interest under the unifying concept of computation. Multiscale modeling and Simulations is a fundamental research topic in CSE exemplifying its interdisciplinary aspects and its unique character as a scientific discipline.

A broad range of scientific and engineering problems involve multi-scale phenomena. The computational study of these problems demands an ever increasing development of novel computational tools and mathematical methods. This Summer Program provided the opportunity to scientists from different scientific fields, faced with multiscaling problems, to exchange ideas and to pursue research in an interdisciplinary fashion. The workshop strengthened existing collaborations and developed new contacts between the participants that we believe will help in defining new scientific frontiers.

The scientific interactions, the weekly lectures and the working sessions showed that scientific cross fertilization can lead to significant advances in the specialized fields as well as in the unifying computational methodologies. The proceedings of this workshop serve as an attestation of this spirit of interdisciplinarity that characterises research in multiscale modeling and simulation.

Zurich,
February 2004

Sabine Attinger
Petros Koumoutsakos

Acknowledgment

We wish to thank all authors for their active participation in the Workshop and their cooperation throughout the completion of this volume.

Professors Parrinello and Kaxiras were instrumental in helping to assemble a distinguished group of participants and in maintaining an exciting scientific environment throughout the workshop.

The workshop would not have been possible without the support of Prosessor Marco Baggiolini and the indispensable help of Dr. Mauro Prevostini from the Universita di Svizzera Italiana.

Financial support for the workshop was provided by the ETHZ Computational Laboratory (CoLab).

Contents

Part I

Mathematical Methods

Some Recent Progress in Multiscale Modeling

Weinan E[1,2], Xiantao Li[2], and Eric Vanden-Eijnden[3]

[1] Mathematic Department, Princeton University, Princeton,NJ 08544
[2] PACM, Princeton University, Princeton, NJ 08544
[3] Courant Institute, New York University, New York, NY 10012

Summary. Some existing multiscale techniques for the modeling of solids and fluids are reviewed. The framework of the heterogeneous multiscale method (HMM) is illustrated with two examples. Finally, time scale issues are discussed.

1 Introduction

In recent years we have seen an explosive growth of activities in multiscale modeling and computation, with applications in many areas including material science, fluid mechanics, chemistry, and biology. It is widely recognized that multiscale techniques will become an essential part of computational science and engineering. The main purpose of the present paper is to review some (not all) of these progresses, with a view toward building systematic, reliable and controlled multiscale techniques for a variety of applications.

The basic set-up will be the following. We are given a system whose microscopic behavior, with state variable u, is described by a given microscopic model. This microscopic model is too inefficient to be used in full detail. On the other hand, we are only interested in the macroscopic behavior of the system, described by the state variable U. U and u are linked together by a compression operator Q:

$$U = Qu.$$

The difficulty stems from the fact that the macroscopic model for U is either not explicitly available or it is invalid in some parts of the computational domain. Our basic strategy is to use the microscopic model as a supplement to provide the necessary information for extracting the macroscale behavior of the system. Our hope is that if this is done properly, the combined macro-micro modeling technique will be much more efficient than solving the full microscopic model in detail [EEN03].

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

2 Examples of Multiscale Methods

For the purpose of a systematic study, it is useful to divide multiscale problems into several different types according to some common features. These features are used in designing efficient modeling strategies. Here we will concentrate on two types of problems [EE03]. Other types of problems can also be conceived, but currently they are less exploited.

2.1 Type A Problems – Dealing with Isolated Defects

Type A problems are problems that have isolated defects near which the macroscopic models are invalid. Elsewhere explicitly given macroscale models are sufficient. Such defects can be shocks and contact lines in fluids, cracks, dislocations and triple junctions in solids or enzymatic reaction regions in biopolymers. Numerous computational techniques have been developed for such problems. An incomplete list includes

1. The nonlocal quasicontinuum method (QC) [TOP96] for simulating isolated defects such as dislocations and cracks in single crystals. In this case the microscale model is molecular mechanics. The macroscale model, which is not explicitly used in QC, is nonlinear elasticity. An adaptive mesh refinement procedure allows QC to identify where the defects are and refine locally to the atomistic scale in order to resolve the full details around the defects.

2. DSMC-gas dynamics [GBCA99]. This is a very elegant extension of the adaptive mesh refinement procedure developed through the years by Bell, Berger, Colella et al. that incorporates the kinetic model at the finest level near shocks or other regions of special interest. The kinetic model is solved using discrete simulation Monte Carlo (DSMC). The continuum model and DSMC are coupled together through exchanges of fluxes at the cell boundaries.

3. MAAD-Macro Atomistic Ab initio Dynamics [ABBK98, ABBK99]. This highly publicized calculation was responsible for inspiring numerous subsequent work on multiscale modeling. Crack propagation in silicon was simulated using a seamless coupling procedure between finite elements away from the crack tip, molecular dynamics around the crack tip and tight-binding at the crack tip. Tight-binding was chosen as the simplest model for describing bond breaking at the crack tip. Molecular dynamics (MD) around the crack tip is capable of modeling processes such as dislocation loop emission. Finally finite element calculation away from the crack tip serves as boundary conditions for MD and at the same time carries the macroscopic deformation.

4. Coupled continuum - MD models for fluids [LLY99, FWF00, OC95, Ha99, NCER]. Numerous attempts have been made to couple continuum models with MD for the modeling of fluids, with potential applications to micro-fluidics. The basic idea is that of heterogeneous domain decomposition: The computational domain is

divided into overlapping regions on which continuum and MD calculations are carried out. The results are matched in the overlapping region to ensure continuity of the overall calculation.

5. Coupled continuum - MD methods for solids [CKBY00, EH02, BX03, WL03]. These are methods that couple continuum models with MD for the simulation of the dynamic behavior of solids. The main issue is the matching condition at the continuum-MD interface. Cai et al. proposed exact boundary condition for linear problems, which provides a conceptual starting point but is too expensive in practice[CKBY00]. E and Huang [EH02] extended the absorbing boundary conditions, traditionally designed for continuum calculations of waves [EM79], to the discrete models of MD. This seems to be a good compromise between exact boundary conditions of Cai et al. and the straightforward extensions of the absorbing boundary conditions for continuum models. At low temperature when no thermal energy is supplied to the system, this method works quite well.

6. Coupled quantum mechanics - molecular dynamics (QM-MM) methods [WL76, ZLY99]. This is a popular method that uses quantum mechanical models in regions of chemical reaction, and classical molecular mechanics models elsewhere.

2.2 Type B Problems – Constitutive Modeling Based on the Microscopic Models

Type B problems are those for which the macroscopic model is not explicitly available, instead it should be indirectly inferred from the underlying microscopic model. Such problems occur in molecular dynamics when the atomistic potential is not explicitly known (which is often the case), or in continuum simulations of solids and fluids when the constitutive relations are not explicitly given. This class of problems are referred to as "equation-free computation", or "solving equations without equations" in [TQK00]. Examples include:

1. Artificial compressibility method for computing incompressible flows [Chr67]. The motivation in this case is to simulate incompressible flows without using the equations for incompressible flows, but rather using the compressible flow equations. Even though the equations for incompressible flows, the Navier-Stokes equations, are explicitly known, solving them numerically is a non-trivial matter. The idea of Chorin's artificial compressibility method is to use instead the compressible flow equations with suitably chosen Mach number. The optimal choice of Mach number is a compromise between accuracy, which favors choosing small Mach number, and computational cost, which favors choosing a larger Mach number. Note that stability constraints the size of the time step in accordance with the Mach number.

2. The Car-Parrinello method [CP85]. One of the key questions in MD simulations is how to get accurate interaction potential between atoms. For many problems the right potential is the Born-Oppenheimer potential obtained by calculating the ground state energy of the electrons when the positions of the nuclei are fixed. However

computing the Born-Oppenheimer potential for realistic systems is a rather daunting task.

In 1985, Car and Parrinello devised a strategy for probing the relevant part of the Born-Oppenheimer surface on the fly during MD simulations. Their method is formulated in the framework of density functional theory. The Lagrangian for the system of nuclei is augmented with new degrees of freedom associated with the Kohn-Sham orbitals with fictitious mass. The new nuclei-orbital system evolves just as any classical system would do, but with energy computed from density functional theory. This method has become a popular tool and has found wide application in material science, chemistry and biology.

The time step size is controlled by the fictitious mass associated with the Kohn-Sham orbitals. On physical grounds, one might be tempted to use electron mass as the fictitious mass, but since we are only interested in the dynamics of the nuclei, it is more efficient to choose a larger fictitious mass so that larger time steps can be used, as long as the desired accuracy for the dynamics of the nuclei is retained. This is in the spirit of the artificial compressibility method [Chr02].

3. Quasi-continuum method (QC) [TOP96, STK00, KO01]. QC can also be used as a way of simulating the macroscopic deformation of a material without using empirical stored energy functional, but using directly the atomic potential. Compared with full atom calculation, the savings come from the fact that the energy for a small cluster of atoms can be used to represent the average energy over a macroscopic cell (Figure 1). Therefore the atoms in the large voids do not have to be visited.

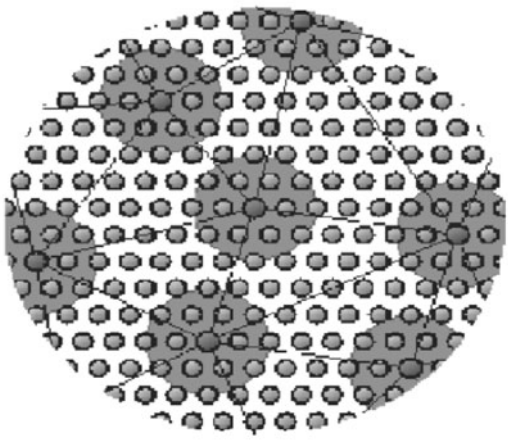


Fig. 1. Schematic illustration of QC (courtesy of M. Ortiz). Only atoms in the small cluster need to be visited during the computation. (See Plate 1 on Page 269)

4. Gas-kinetic scheme [Xu94]. Even though originally it was not designed for the purpose of multiscale modeling, the gas-kinetic scheme provides one of the best examples of multiscale methods. Here the motivation is to compute the macroscopic behavior of gases without using Euler's equation of gas dynamics but using directly the kinetic equations. Starting from a finite volume framework, hydrodynamic fluxes at cell boundaries are computed locally near the cell boundaries at each macro time step from the kinetic models, using a reconstruction-evolution-averaging procedure. This scheme served as the starting point for the heterogeneous multiscale method discussed later.

5. Stochastic ODEs [VE03]. This is discussed in Section 4 where we concentrate on time-scale issues.

6. Gap-Tooth scheme [TQK00]. Gap-Tooth scheme [TQK00] is an attempt to extend the philosophy illustrated in Figure 1 to a larger class of problems. Starting with a macroscopic initial data, the microscopic model is solved to the new macroscopic time step in small boxes (the tooth) centered around macroscale grid points, with suitable boundary conditions that are obtained from the macroscale data. The microscale solutions are then averaged in the boxes, and the macro state at the new time step is obtained by interpolating these averaged data. Gap-tooth scheme has been successfully applied to the one-dimensional heat equation, using the heat equation itself as the microscale model [TQK00]. More recently it has been extended to solving the one-dimensional parabolic periodic homogenization problem [Giov]. However, it seems that this method is not suited for problems for which the macroscale dynamics is hyperbolic, e.g. convection equations, inviscid Burgers equations or equations of gas dynamics.

In connection with the quasi-continuum method, the clusters at the centers of the elements in Figure 1 would be the teeth, the voids would be the gaps.

2.3 Other Methods

We discuss briefly several other methods that are relevant to multiscale modeling.

1. Coarse-grained MD (CGMD) [RB99, RB98]. This procedure works as follows. Starting with the microscopic Hamiltonian and after defining a set of coarse-grained degrees of freedom, one obtains a reduced Hamiltonian for the coarse-grained variables by averaging out the microscopic Hamiltonian with respect to the conditional equilibrium distribution. The coarse-grained MD is defined with respect to this reduced Hamiltonian.

2. Optimal prediction [Chr98, Chr02]. In many complex systems we often have the problem that either the model or part of the data are not explicitly given. The idea of optimal prediction is to look for the best approximate solution with given information on the model and the data. A key technical aspect of this scheme is the norm used in defining the best approximation. In [Chr98, Chr02] Chorin et al. use the L^2 norm

with respect to the equilibrium distribution. This requires knowing some information about the equilibrium distribution which can then be obtained perturbatively.

In general this procedure may result in an effective dynamics that is memory-dependent, as in the Zwanzig-Mori procedure. But the procedure simplifies if the problem has certain special features such as scale separation.

This philosophy is quite general. Indeed, coarse-grained MD, for example, can be viewed as a special case of optimal prediction.

3. Projective dynamics for kinetic Monte Carlo models [KNR98, Nov01] . This is a procedure for accelerating kinetic Monte Carlo simulations that involve disparate rates. The idea is to reduce the dynamics to a Markov chain that involves only the slow rates, by projecting out the fast processes.

4. Coarse-grained Monte Carlo [KMV1, KMV2]. This is a semi-analytical and semi-numerical multiscale technique that systematically coarse grains Monte Carlo models with long range interaction [KMV1]. Interaction parameters for the coarse-grained model were calculated analytically by suitably averaging the parameters in the microscopic model. Detailed balance is enforced at the coarse-grained level. Self-consistency is guaranteed by the fact that the microscopic model and the coarse-grained model share the same mesoscopic limit.

2.4 The Heterogeneous Multiscale Method

The heterogeneous multiscale method (HMM) [EE03] is an attempt to construct a unified framework for designing efficient simulation methods that couple the macro and microscale models. The framework should apply for both type A and type B problems. The basic principle of HMM is that one should start with a macroscale solver, taking into account as much as possible what is known about the macroscale process, and use the microscale model to provide the missing macroscale data that are necessary for implementing the macroscale solver. When measuring the macroscale data, the microscale model should be constrained by the (local) macro state of the system.

One distinct feature of HMM is its emphasis on selecting the right macroscale solver. This is especially important for problems whose macroscale behavior contains singularities. In general the principles for selecting the macroscale solver are the following: (1) It should be stable. (2) It should facilitate coupling with the microscale model.

3 Selecting the Macroscale Solvers

The emphasis on selecting the appropriate macro solver is motivated by the fact that for many problems of interest, the macroscale processes are also quite complicated and they may contain singularities, instabilities or phase transitions. The numerical problems can be very non-trivial even if a valid macroscale model is explicitly available.

To illustrate the selection of macro solvers, we will discuss two examples from gas dynamics. The first example is continuum gas dynamics locally corrected near shocks by the kinetic model. This is a type A problem. For this problem, the macroscopic process is gas dynamics, the microscopic process is described by a kinetic model. Therefore we will choose the kinetic scheme for gas dynamics.

Recall the Euler equations for ideal gases,

$$\begin{cases} \mathrm{d}\rho_t + \nabla \cdot (\rho \boldsymbol{u}) = 0, \\ \mathrm{d}(\rho \boldsymbol{u})_t + \nabla \cdot (\rho \boldsymbol{u} \otimes \boldsymbol{u} + \rho\theta I) = 0, \\ \mathrm{d}E_t + \nabla \cdot (\rho \boldsymbol{u}(|\boldsymbol{u}|^2 + \frac{5}{2}\theta)) = 0, \\ E = \rho(\frac{1}{2}|\boldsymbol{u}|^2 + \frac{3}{2}\theta), \end{cases} \tag{1}$$

where $\theta = RT$ with T being the temperature. A more detailed model is the Boltzmann equation,

$$f_t + \boldsymbol{\xi} \cdot \nabla f = \frac{1}{\varepsilon} C(f). \tag{2}$$

The function $f(\boldsymbol{x}, \boldsymbol{\xi}, t)$ is the one particle phase-space distribution function. The variable $\boldsymbol{\xi}$ dictates velocity distribution. When the mean free path $\varepsilon \ll 1$, the system is close to local equilibrium states, or the local Maxwellians,

$$M(\boldsymbol{x}, \boldsymbol{\xi}, t) = \frac{\rho(\boldsymbol{x}, t)}{(2\pi\theta(\boldsymbol{x}, t))^{3/2}} \exp\Big(-\frac{(\boldsymbol{\xi} - \boldsymbol{u}(\boldsymbol{x}, t))^2}{2\theta(\boldsymbol{x}, t)}\Big). \tag{3}$$

with θ being the absolute temperature. In this regime, (1) can be derived using Chapman-Enskog expansion and moment closure with

$$\rho = \int f d\boldsymbol{\xi}, \quad \rho \mathbf{u} = \int f \boldsymbol{\xi} d\boldsymbol{\xi}, \quad E = \int f \frac{|\boldsymbol{\xi}|^2}{2} d\boldsymbol{\xi}. \tag{4}$$

Since we are interested in coupling the macroscale solver with the kinetic equation around the shock, we will choose a kinetic scheme as the macroscopic solver. Kinetic schemes take advantage of the fact that the solution of the Boltzmann equation is close to the local Maxwellian. We denote the grid points for the macro-mesh by x_j, the cell interface by $x_{j+1/2}$ and the time steps by t^n. For simplicity we only present the one dimensional version of the method. Extension to higher dimension is straightforward. For first order method, we represent the solution as piece-wise constant, i.e.

$$(\rho, \rho u, E) = (\rho_j, \rho_j u_j, E_j), \quad x \in (x_{j-1/2}, x_{j+1/2}],$$

and the grid values are the cell averages of the corresponding variables.

The derivation of the kinetic schemes consists of two steps. In the first step, we 'solve' the transport equation,

$$f_t + \xi f_x = 0.$$

Specifically we multiply the transport equation by $\xi, \xi^2, \xi^3/2$ and integrate within one grid cell and one time step, and also over the velocity space, to get

$$\begin{cases} \mathrm{d}\rho_j^{n+1} - \rho_j^n + \frac{\Delta t}{\Delta x}\big(F^{(1)}_{j+1/2} - F^{(1)}_{j-1/2}\big) = 0, \\ \mathrm{d}(\rho u)_j^{n+1} - (\rho u)_j^n + \frac{\Delta t}{\Delta x}\big(F^{(2)}_{j+1/2} - F^{(2)}_{j-1/2}\big) = 0, \\ \mathrm{d}E_j^{n+1} - E_j^n + \frac{\Delta t}{\Delta x}\big(F^{(3)}_{j+1/2} - F^{(3)}_{j-1/2}\big) = 0. \end{cases} \tag{5}$$

where

$$\rho_j^n = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \rho(x, t^n)dx,$$

is the average of the solution within one cell, and

$$\boldsymbol{F}_{j+1/2} = \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} \int_{\mathbb{R}} f(x_{j+1/2}, \xi, t) \begin{pmatrix} \xi \\ \xi^2 \\ \frac{1}{2}\xi^3 \end{pmatrix} d\xi dt$$

is the numerical fluxes across the cell interfaces.

The distribution function f can be decomposed into two parts: a left-moving part and a right-moving part. This gives rise to a natural splitting of the numerical fluxes:

$$\boldsymbol{F} = \boldsymbol{F}^+ + \boldsymbol{F}^-, \quad \text{with } \boldsymbol{F}^\pm = \int_{\mathbb{R}^\pm} f(x^\mp_{j+1/2}, \xi, t) \begin{pmatrix} \xi \\ \xi^2 \\ \frac{1}{2}\xi^3 \end{pmatrix} d\xi. \tag{6}$$

Assume that $f(x, \xi, t)$ can be approximated by $M(x - \xi t, \xi, t^n)$. By direct computation we get,

$$\boldsymbol{F}^\pm = \begin{pmatrix} \mathrm{d}\rho u A^\pm(S) \pm \frac{\rho}{2\sqrt{\pi\beta}} B(S) \\ \mathrm{d}(p + \rho u^2)A^\pm(S) \pm \frac{\rho u}{2\sqrt{\pi\beta}} B(S) \\ \mathrm{d}(pu + \rho u e)A^\pm(S) \pm \frac{\rho}{2\sqrt{\pi\beta}}(\frac{p}{2\rho} + e)B(S) \end{pmatrix} \tag{7}$$

where

$$A^\pm = \frac{1 + \mathrm{erf}(S)}{2}, \quad B(S) = e^{-S^2}, \quad S = \frac{u}{\sqrt{2RT}}, \quad p = \rho RT.$$

In the second step, we include the collision term,

$$f_t = \frac{1}{\varepsilon} C(f).$$

Since the solution is close to the local Maxwellian, we include the effect of collision by simply setting $f(x, \xi, t^{n+1}) = M(x, \xi, t^{n+1})$. Also by doing this, one can continue to compute the numerical flux as in the previous time step. Despite the fact that the scheme is derived from kinetic models, the method only involves the computation of the numerical fluxes, and operates in the physical space. To achieve second

order accuracy, one can use piecewise linear representation of the solutions. For other types of kinetic schemes see [P92, Xu94].

In regions of large gradients where the local equilibrium approximation is inaccurate, we use directly the kinetic model. The transport part of the kinetic equation is solved using an upwind scheme:

$$vf_{j+1/2} = v^+ f_j + v^- f_{j+1} \tag{8}$$

with $v^+ = \frac{1}{2}(v + |v|)$ and $v^- = \frac{1}{2}(v - |v|)$. The collision part is added afterwards, which is usually done by particle method. The coupling of these two models in different regions is done as follows. Suppose that the interface between the kinetic region and the fluid domain is at $x_{k+1/2}$, with the fluid region on the left and the kinetic region on the right. Since we are using upwind scheme to solve the transport part of the Boltzmann equation, we only need the boundary condition for $\xi > 0$. For this we impose

$$f(x, \xi, t) = M(x^-_{k+1/2}, \xi, t), \quad \xi > 0$$

where the right hand side is the local Maxwellian from the fluid side. On the other hand, in order to update the fluid equations to the next time step, we need to compute the fluxes at the interface. The splitting of the fluxes will automatically incorporate the contribution from the kinetic domain,

$$\begin{aligned} \boldsymbol{F}_{k+1/2} = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} dt \Big\{ & \int_{\mathbb{R}^+} M(x^-_{k+1/2}, \xi, t) \begin{pmatrix} \xi \\ \xi^2 \\ \frac{1}{2}\xi^3 \end{pmatrix} d\xi \\ & + \int_{\mathbb{R}^-} f(x^+_{k+1/2}, \xi, t) \begin{pmatrix} \xi \\ \xi^2 \\ \frac{1}{2}\xi^3 \end{pmatrix} d\xi \Big\}. \end{aligned} \tag{9}$$

The coupling procedure will have to be more complicated if other macroscopic solvers were used. With all the needed components ready, we can write the computational procedure within one loop for$t \in [t^n, t^{n+1})$:

1. In the continuum region, with $(\rho, \rho \boldsymbol{u}, E)$ known at t^n, compute the macroscopic fluxes using (7) and (9), and evolve the macroscopic variables using the kinetic scheme(5);
2. In the kinetic region, compute the fluxes using (8) to evolve the kinetic equation and then include the collision effect by e.g. Monte Carlo method;
3. Measure the gradient of the macroscale variables to determine the new kinetic and continuum regions;
4. For cells that change character, say from continuum to kinetic, reconstruct the microstate using local Maxwellian; Otherwise reduce the microscale distribution to continuum variables by integrating out the phase space variable.

This procedure is repeated until the desired computational goal is reached.

Our second example is aiming at more general continuum models at macroscale including fluid and solid mechanics, and this time we assume that the equation of

state is not given to us, instead we need to extract it from an underlying atomistic model of molecular dynamics. The macro solver in this case should be a conservative shock capturing scheme. Such schemes can be divided into two types, depending on whether a characteristic decomposition is used [LeV90]. Since the nonlinear flux functions are not explicitly given to us, it is much easier to use a central type of scheme that does not use characteristic decomposition, the simplest of which is the Lax-Friedrichs scheme. Comparing with the Godunov scheme, central scheme uses the flux function at the cell center, which makes the coupling with MD simulation much easier.

The macroscopic equations are the usual conservation law of density (deformation for solids), momentum and energy. We use the generic form of conservation laws in one dimension:

$$\boldsymbol{u}_t + \boldsymbol{f}(\boldsymbol{u})_x = 0. \tag{10}$$

The first order central scheme [NT90] constructs the solutions as piece-wise constants, which are the averaged values over each cell:

$$\boldsymbol{u}_j^n = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \boldsymbol{u}(x, t^n)dx.$$

Integration over $[x_j,\ x_{j+1}] \times [t^n,\ t^{n+1})$ leads to the following scheme:

$$\boldsymbol{u}_{j+1/2}^{n+1} = \frac{\boldsymbol{u}_j^n + \boldsymbol{u}_{j+1}^n}{2} - \frac{\Delta t}{\Delta x}\big(\boldsymbol{f}(\boldsymbol{u}_{j+1}^n) - \boldsymbol{f}(\boldsymbol{u}_j^n)\big). \tag{11}$$

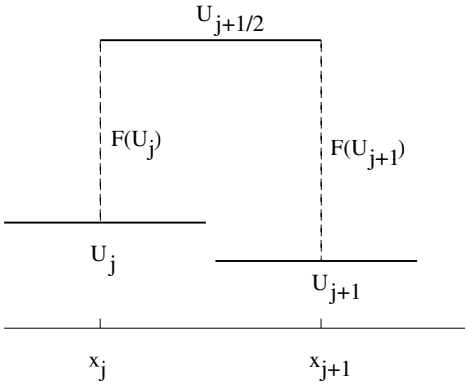


Fig. 2. Central scheme: starting with piecewise constant solution, one computes fluxes at x_j and x_{j+1}, and integrates the conservation laws to the next time step, where the grid points are shifted to the midpoints $x_{j+1/2}$.

At time t^{n+1} the numerical solutions are defined on a different set of grid points, as show in Figure 2. By applying the procedure again, one returns back to the original

grid. The use of staggered grids has been shown to have less numerical viscosity than the usual Lax-Friedrichs scheme. Time steps have to be chosen to ensure stability. Let $A = \nabla \boldsymbol{f}$ and λ be the largest among the absolute values of the eigenvalues for A. Then it can be shown [NT90] that the central scheme is stable if

$$\lambda \frac{\Delta t}{\Delta x} \leq \frac{1}{2}.$$

In our model, the fluxes are computed from local MD simulations: using the macroscale variables $\mathbf{u}$ we invoke the molecular system. For example, one can first arrange the atom positions according to some lattice structure corresponding to local density or deformation, and compute the potential energy afterwards. Subtracting the kinetic and potential energy from the total energy gives temperature. The velocity can thus be initialized by Gaussian distribution with the right mean velocity and temperature. We then evolve the system to local equilibrium, and estimate the fluxes via time and ensemble averaging. HMM stresses that the initialization and evolution of the molecular system should be consistent with the macroscale states. This is discussed in full detail in [LE03], including how to maintain deformation and temperature gradients. Given macroscale variables $\boldsymbol{u}_j$ at time t^n the overall algorithm goes as follows:

1. Initialize the MD system at the cell centers;
2. Evolve the MD system using appropriate boundary conditions;
3. As the system equilibrates, estimate the needed fluxes via time/ensemble averaging;
4. Use the conservative scheme (11) to update the macro variables to next time step;

For a discussion of the MD boundary condition, see [LE03].

4 Time-Scale Problems in ODEs and Stochastic ODEs

We will distinguish four different types of multiple time-scale problems.

1. The first type, perhaps the simplest, is associated with the existence of a slow manifold, onto which the system is very rapidly attracted to. An example of this kind is the system

$$\begin{cases} \mathrm{d}\dot{x} = -\frac{1}{\varepsilon}(x - f(y)), \\ \dot{y} = g(x, y). \end{cases} \tag{12}$$

In this example the slow manifold is given by the graph $\{(x, y) : x = f(y)\}$. Starting from any initial condition, the system is attracted to this manifold in $O(\varepsilon)$ time-scale. x is the fast variable and y is the slow variable. On the slow manifold, y evolves according to an effective equation

$$\dot{y} = G(y) = g(f(y), y). \tag{13}$$

This class of problems are discussed in the context of stiff ODEs.

2. The second class of problems are often discussed in the context of averaging methods. In these problems, the slow variables feel the average effect of the fast variables. One simplest example is

$$\begin{cases} \mathrm{d}\dot{\varphi} = \frac{1}{\varepsilon}\omega(I) + f(\varphi, I) \\ \dot{I} = g(\varphi, I) \end{cases} \tag{14}$$

where f and g are periodic functions of φ, here the fast variable. The averaged system for the slow variable I is given by

$$\dot{I} = \frac{1}{2\pi}\int_0^{2\pi} g(\varphi, I)d\varphi = G(I). \tag{15}$$

Both (12) and (14) are special cases of

$$\begin{cases} \mathrm{d}\dot{x} = \frac{1}{\varepsilon}f(x, y), \\ \dot{y} = g(x, y). \end{cases} \tag{16}$$

Assume that for fixed y, the dynamics given by equation $\dot{x} = f(x, y)/\varepsilon$ is ergodic and denote by $\mu_y(dx)$ the unique invariant measure for x. The effective dynamics for the slow variable y is then given by

$$\dot{y} = G(y) \qquad \text{with } G(y) = \int g(x, y)\mu_y(dx) \tag{17}$$

(12) corresponds to the case when $\mu_y(dx) = \delta(x - f(y))$ and (13) corresponds to the case when $\mu_I(d\varphi) =$ Lebesgue measure on the circle. The first class of problems are characterized by the fact that μ is dirac measure. (16) can be far more general. In particular, it can be a system of stochastic ODEs.

3. Effective stochastic dynamics. In these problems, the fast scale dynamics act effectively as noise on the slow dynamics. This is believed to be the case for example in molecular dynamics, which after coarse graining, may result in Brownian dynamics. Consider (16) in a special situation where the equation $\dot{x} = f(x, y)$ is ergodic but

$$\int g(x, y)\mu_y(dx) = 0. \tag{18}$$

In this case, the dynamics on the $O(1)$ time-scale is trivial, $\dot{y} = 0$, and the evolution of the slow variable arises on the longer, $O(\varepsilon^{-1})$ time-scale. On this time-scale, fluctuations becomes important, and the effective equation for the slow variable is an SDE:

$$\dot{y} = b(y) + \sigma(y)\dot{W}. \tag{19}$$

Here $\dot{W}$ is a white-noise process, and the coefficients $b(y)$ and $\sigma(y)$ satisfy

$$\begin{cases} \mathrm{d}b(y) = \int_0^\infty \int (g(x, y) \cdot \nabla_x)g(\varphi_t(x), y)\mu_y(dx)dt, \\ \mathrm{d}\sigma(y)\sigma^T(y) = \int_0^\infty \int g(x, y)g^T(\varphi_t(x), y)\mu_y(dx)dt, \end{cases} \tag{20}$$

where $\varphi_t(x)$ denotes the solution of $\dot{x} = f(x, y)$ at fixed y with initial condition $\varphi_0(x) = x$. Besides ergodicity, (19) requires enough mixing for the dynamics of the fast variable x in order that the time-integrals in (20) be finite. When the dynamics of the fast variables is governed by an SDE, this criterion is usually met, but explicit examples when y satisfies an ODE are rare, see for example [Bun00].

It is also interesting to note that even if the expectation in (18) is non-zero, the effective equation (17) may only be valid on finite time intervals, and the small random perturbations that arise on the $O(\varepsilon^{-1})$ time scale are essential to describe the long-time behavior of the system. Here is an example. Consider

$$\begin{cases} \mathrm{d}\dot{x} = -\frac{1}{\epsilon}(y - x) + \frac{1}{\sqrt{\epsilon}}\dot{W} \\ \dot{y} = x - y^3. \end{cases} \tag{21}$$

For this system the effective equation (17) reads

$$\dot{y} = y - y^3, \tag{22}$$

according to which the motion of y converges to the stable fixed points $+1$ or -1 depending on the sign of the initial y. However, accounting for the next order correction in ε as in (19), the effective equation for y becomes

$$\dot{y} = y - y^3 + \sqrt{\epsilon}\,\dot{W}. \tag{23}$$

The additional noise term in this equation eventually drives the system back and forth between the two positions $y = \pm 1$. Note that this happens on the Arrhenius time-scale which is $O(\exp(1/\epsilon))$ because the switching from -1 to $+1$ involves a barrier crossing event. In this case, the effective equation (22) can be further reduced to that of a two-state Markov chain on $\{-1, +1\}$. This type of effective dynamics is discussed next.

4. Activated dynamics. In these examples the system spends most of its time in confined metastable sets, but occasionally hops from one metastable state to another. A simple example is (23) which can be rewritten as

$$\frac{dy}{dt} = -\frac{\partial V}{\partial y}(y) + \sqrt{\varepsilon}\,\dot{W}$$

where $V(y) = \frac{1}{4}(1 - y^2)^2$. Here the metastable states are around $y = \pm 1$. These examples are characterized by the exponentially long (in ε) time-scales that separate the individual hopping (transition) events between different metastable states.

Different techniques are required for these different problems. Stiff solvers are developed for problems of the first kind. These solvers are often implicit. But the work of Lebedev et al. makes an important exception. Lebedev et al. developed explicit Runge-Kutta methods, whose stage parameters are chosen to optimize their stability regions according to certain given distribution of the spectrum, which are assumed to be close to the real axis. For the special case of (12), the spectrum has a

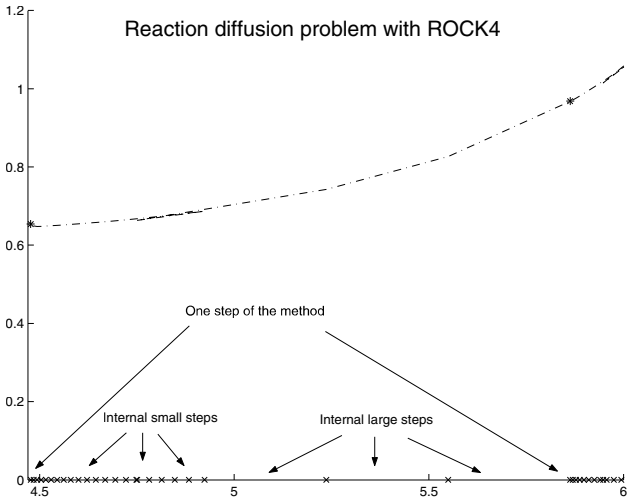


Fig. 3. Multistage Runge-Kutta: each step of the method consists of many small stages followed by a few big stages. Horizontal axis shows the time steps and internal stages within a time step. Vertical axis shows the computed solutions in a reaction-diffusion problem. (Courtesy of A. Abdulle)

large gap. In this case, each step of the optimized Runge-Kutta method consists of many small stages followed by a few big stages, as shown in Figure 3.

In Lebedev's work only second order methods were discussed. Fourth order accurate Chebychev methods were constructed in [Ab02].

The recent work of Gear et al. [GK03] can be considered as a simplified version of Chebychev methods that also uses a combination of small and large time steps but is otherwise unoptimized, in contrast to the Chebychev methods.

Chebyshev methods do not apply when the spectrum of the problem is far away from the real axis, as in the second kind of problems discussed above. However, these problems can be readily treated using HMM [E03, ET]. The accuracy of HMM for these problems are greatly affected by the averaging procedure in the force estimator in HMM. These issues are discussed in [E03, ELV, ET] (see also [FV]).

For stochastic ODEs, [VE03] proposed methods that use a combination of macro- and micro-solvers. The basic idea is to estimate the drift and diffusion coefficients in the macroscopic dynamics directly from solutions of the microscale model. Further analysis of these methods are presented in [ELV].

Even though Brownian dynamics has been used as a standard tool in the modeling of complex fluids, at the present time, there still lacks a systematic coarse graining procedure for producing Brownian dynamics from more detailed models such as

molecular dynamics. This is an area that deserves much more attention in the near future.

Activated dynamics can be considered as a sequence of transition events between metastable regions in the configuration space. Each event has an initial state and a final state. There has been two different approaches to activated dynamics, depending on whether the final state in the transition is assumed to be known. If the final state in a transition event is known, then the only remaining issue is to find the transition pathways and transition rates. Numerous numerical strategies have been proposed for this purpose. We refer to the papers [EV, Jon] in this volume for discussions on this class of methods. Several methods have been proposed to deal with the case when the final state is not assumed to be known. The best known among this class of methods are the hyperdynamics [Vot97] and metadynamics [LP02, ILP03]. Hyperdynamics follows the dynamics in a modified potential energy surface and recovers information for the original dynamics using transition state theory. A related work is TAD (temperature accelerated dynamics) [SV00]. Metadynamics proceeds by first constructing the free energy surface associated with some judiciously chosen reaction coordinates, and then sample efficiently that free energy surface. Both the hyperdynamics and metadynamics are very promising ideas that should be exploited in the future.

5 Conclusions and Outlook

To conclude, let us ask the question: What is the current status of multiscale modeling? One thing that can be said about it is that it is now very popular. There are popular methods, such as the Car-Parrinello method and more recently the quasicontinuum method. There are popular ideas, such as hybrid atomistic-continuum methods, and macroscopic methods that use input obtained from pre-computed results using the microscopic models. There are also commonly realized issues, such as matching conditions between different levels of physical models.

As is the case for any fashionable subject, there is also the need for processing the rapidly generated information in this area with great care. With few exceptions, most of the recently proposed multiscale modeling techniques have not been tested on truly challenging problem. Most work remains at the level of testing concepts. Few have gone through the rigorous procedure of verifying the results. Instead, most seem to be content with producing reasonably looking pictures. Analytical results on error control are very rare. Moreover, some of the proposed multiscale methods are more expensive than solving directly the original microscale problem (see for example the discussions in [MY]).

From a mathematical perspective, it is desirable to have a unified framework for studying multiscale methods. So far HMM seems to be a reasonable candidate. Several classes of problems are solved using this framework, including atomistic-continuum modeling for the dynamics of solids [LE03] and fluids [Ren], interface dynamics [Ch03], and homogenization problems [AE03, MY]. In addition, several class of methods, including HMM for the homogenization problems [AE03, EMZ,

MZ], ODEs [E03, ET], multiscale method for stochastic ODEs [VE03, ELV], and the quasi-continuum methods [EM03], have been analyzed using this framework. But it should be emphasized that HMM provides only a starting point, for most problems of interest, highly non-trivial practical issues still remain. While HMM facilitates macro-micro coupling, it does not solve all the problems of macro-micro coupling. This is illustrated clearly in [Ren, LE03]. Solutions to the remaining coupling issues are useful not only in the context of HMM, but also for other coupling strategies such as sequential coupling.

Finally it should be mentioned that most multiscale modeling work deals with almost homogeneous systems such as fluids. There is relatively little work for strongly heterogeneous systems such as macromolecules. The latter will undoubtedly be an area of great interest in the near future.

Acknowledgement. We are grateful to A. Abdulle and M. Ortiz for providing Figure 3 and Figure 1 respectively. The work of E is supported in part by ONR grant N00014-01-1-0674. The work of X.T. Li is supported in part by ONR grant N00014-01-1-0674. The work of E. Vanden-Eijnden is supported in part by NSF grants DMS01-01439, DMS02-09959 and DMS02-39625..

References

[ABBK98] F.F. Abraham, J.Q. Broughton, N. Bernstein and E. Kaxiras, "Spanning the continuum to quantum length scales in a dynamic simulation of brittle fracture, *Europhys. Lett.*, 44 (6), 783–787 (1998).

[ABBK99] F.F. Abraham, J.Q. Broughton, N. Bernstein and E. Kaxiras, "Concurrent coupling of length scales: Methodology and application," *Phys. Rev. B* **60** (4), 2391–2402 (1999).

[Ab02] A. Abdulle, "Fourth order Chebychev methods with recurrence relations", *SIAM J. Sci. Comput.* **23**, 2041–2054 (2002).

[AE03] A. Abdulle and W. E, "Finite difference HMM for homogenization problems", *J. Comput. Phys.* **191**, 18–39 (2003).

[Bou96] J. Bourgat, P. Le Tallec and M. Tidriri, "Coupling Boltzmann and Navier-Stokes equations by friction", *J. Comput. Phys.* **127**, 227 (1996).

[Br92] A. Brandt, "Multigrid methods in lattice field computations", *Nuclear Physics B Proc. Suppl.* **26** 137–180 (1992).

[BX03] T. Belytschko and S. Xiao, "Coupling methods for continuum model with molecular model", International Journal for Multiscale Computational Engineering, 1(1), 115-126, 2003.

[Br02] A. Brandt, "Multiscale scientific computation: Review 2001". *Lecture Notes in Computational Science and Engineering*, T.J. Barth et.al eds., Springer, 2002.

[Bun00] L. A. Bunimovich, "Existence of transport coefficients", in *Hard Ball Systems and the Lorentz Gas*, D. Szasz ed., Springer, 2000.

[CP85] R. Car and M. Parrinello, " Unified approach for molecular dynamics and density-functional theory", *Physical Review Letters* **55**, 2471–2474 (1985).

[CKBY00] W. Cai, M. de Koning, V.V. Bulatov and S. Yip, "Minimizing boundary reflections in coupled-domain simulations," *Phys. Rev. Lett.* **85**, 3213–3216 (2000).

[Ch03] L. T. Cheng and W. E, "The heterogeneous multiscale method for interface dynamics", *Contemporary Mathematics:* **330**, 43–53 (2003).

[Chr67] A. J. Chorin, "A numerical method for solving incompressible viscous flow problems", *J. Comput. Phys.* **2**, 12–26, 1967.

[Chr98] A. J. Chorin, A. Kast and R. Kupferman, "Optimal prediction of underresolved dynamics", *Proc. Nat. Acad. Sc. USA* **95**, 4094–4098 (1998).

[Chr02] A. J. Chorin, O. Hold and R. Kupferman, "Optimal prediction with memory", *Physica D* **166**, 239–257 (2002).

[Dsh86] S. M. Deshpande, Kinetic theory based new upwind methods for invisid compressible flows, *AIAA paper*, No. 86-0275.

[E03] W. E, "Analysis of the heterogeneous multiscale method for ordinary differential equations", *Comm. Math. Sci.* **1**, 423–436 (2003).

[EE03] W. E and B. Engquist,"The heterogeneous multi-scale methods", *Comm. Math. Sci*. **1**, 87–132 (2003).

[EEN03] W. E and B. Engquist, "Multiscale modeling and computation", *Notices of the American Mathematical Society* **50**, 1062–1070 (2003).

[EEH03] W. E, B. Engquist and Z. Huang, "Heterogeneous multi-scale method – a general methodology for multi-scale modeling", *Phys. Rev. B* **67**, 092101 (2003)

[EH01] W. E and Z. Huang, "Matching conditions in atomistic-continuum modeling of materials," *Phys. Rev. Lett.* **87**, 135501 (2001).

[EH02] W. E and Z. Huang, "A dynamic atomistic-continuum method for the simulation of crystalline materials", *J. Comput. Phys.* **182**, 234–261 (2002).

[ELV] W. E, D. Liu and E. Vanden-Eijnden, "Analysis of multiscale techniques for stochastic dynamical systems", preprint, 2003.

[EM03] W. E and P. B. Ming, "Analysis of the quasi-continuum method, I", preprint, 2003.

[EV] W. E and E. Vanden-Eijnden, "Conform dynamics and transition pathways in complex system", this volume.

[EMZ] W. E, P. B. Ming and P. W. Zhang, "Analysis of the heterogeneous multiscale method for elliptic homogenization problems", preprint. *Available at http://www.math.princeton.edu/multiscale.*

[EM79] B. Engquist and A. Majda, "Radiation boundary conditions for acoustic and elastic wave calculations," *Comm. Pure Appl. Math.* **32**, 313–357 (1979).

[ET] B. Engquist and R. Tsai, "The Heterogeneous Multiscale Methods for a class of Stiff ODEs", submitted.

[FV] I. Fatkullin and E. Vanden-Eijnden, "Effective dynamics and efficient computations. Applications to Lorenz 95 system", submitted.

[FWF00] E. G. Flekkoy, G. Wagner and J. Feder, *Europhys. Lett.* **52**, 271 (2000).

[GBCA99] A. L. Garcia, J. B. Bell, W. Y. Crutchfield and B. J. Alder, "Adaptive mesh adn algorithm refinement using direct simulation Monte Carlo" *J. Comput. Phys.* **154**, 134–155 (1999).

[GK03] C.W. Gear and I.G. Kevrekidis, "Projective methods for stiff differential equations: problems with gaps in their eigenvalue spectrum," *SIAM J. Sci. Comp.* **24**, 1091-1106, 2003.

[Ha99] N. G. Hadjicontantinou, "Hybrid atomistic-continuum formulations and the moving contact line problem", *J. Comput. Phys.* **154**, 245–265 (1999).

[HW91] E. Hairer and G. Wanner, *"Solving Ordinary Differential Equations II, Stiff and Differential-Algebraic Problems"*, Springer-Verlag, 1991

[ILP03] M. Iannuzzi, A. Laio, and M. Parrinello, "Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics,", *Phys. Rev. Lett.* **90**, 238302 (2003)

[Jon] H. Jonsson, this volume.

[KMV1] M. Katsoulakis, A. J. Majda and D. G. Vlachos, "Coarse-grained stochastic processes for lattice systems", *Proc. Natl. Acad. Sci. USA* **100**, 782–787 (2003).

[KMV2] M. Katsoulakis, A. J. Majda and D. G. Vlachos, Coarse-grained stochastic processes and Monte Carlo simulations in lattice systems, *J. Comp. Phys.* **186**, 250-278 (2003).

[TQK00] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg and C. Theodoropoulos, "Equation-free multiscale computation: enabling microscopic simulations to perform system-level tasks", in preparation.

[KO01] J. Knap and M. Ortiz, "An analysis of the quasi-continumm method", *J. Mech. Phys. Sol.*, **49**: 1899–1923 (2001).

[KNR98] M. Kolesik, M. A. Novotny and P. Rikvold, "Projection method for statics and dynamics of lattice spin systems", *Phys. Rev. Lett.* **80**, 3384–3387 (1998).

[LP02] A. Laio and M. Parrinello, "Escaping free-energy minima," *Proc. Nat. Acad. Sc. USA* **99**, 12562–12566 (2002).

[LF76] V. I. Lebedev and S. I. Finogenov, "Explicit methods of second order for the solution of stiff systems of ordinary differential equations", *Zh. Vychisl. Mat. Mat Fiziki*, **16**. 895–910 (1976).

[LeV90] R. LeVeque, *"Numerical Methods for Conservation Laws,"* Birkhäuser, 1990.

[LLY99] J. Li, D. Liao and S. Yip, "Nearly exact solution for coupled continuum/MD fluid simulation", *J. Computer-Aided Materials Design* **6**, 95–102 (1999).

[LE03] X.T. Li and W. E, "Heterogeneous multiscale method for the modeling of solids", preprint, 2003.

[MT] R. E. Miller and E. B. Tadmor, "The quasicontinuum method: Overview, applications and current directions", preprint.

[MY] P. B. Ming and X. Yue, "Numerical methods for multiscale elliptic problems", preprint. *Available at http://www.math.princeton.edu/multiscale.*

[MZ] P. B. Ming and P. W. Zhang, "Analysis of the heterogeneous multiscale method for dynamic homogenization problems", preprint.

[NT90] H. Nessyahu and E. Tadmor, Nonoscillatory central differencing for hyperbolic conservation laws, *J. Comp. Phys.* **87**, 408–463 (1990).

[Nov01] M. A. Novotny, "A tutorial on advanced dynamic Monte Carlo methods for systems with discrete state spaces", *Ann. Rev. Comput. Phys.,* 153–210 (2001).

[OC95] S. T. O'Connell and P. A. Thompson, *Phys. Rev. E* **52**, R5792 (1995).

[NCER] X. Nie, S. Chen, W. E and M. O. Robbins, submitted to *Phys. Fluids.*

[P92] B. Perthame, Second-order Boltzmann schemes for compressible Euler equations in one and two space dimensions. *SIAM J. Numer. Anal.* **29**, 1–19 (1992).

[Ren] W. Ren and W. E, "Heterogeneous multiscale method for the modeling of fluids", in preparation.

[RB98] R.E. Rudd and J.Q. Broughton, "Coarse-grained molecular dynamics and the atomic limit of finite elements," *Phys. Rev. B* **58**, R5893–R5896 (1998).

[RB99] R.E. Rudd and J.Q. Broughton, "Atomistic simulation of MEMS resonators through the coupling of length scales," *J. Modeling and Simulation of Microsystems* **1**, 29–38 (1999).

[Giov] G. Samaey, private communication, 2003.

[SSE03] T. Schulze, P. Smereka and W. E, "Coupling kinetic Monte-Carlo with continuum models with applications to epitaxial growth", *J. Comput. Phys.* **189**, 197–211 (2003).

[STK00] G.S. Smith, E.B. Tadmor, and E. Kaxiras, "Multiscale simulation of loading and electrical resistance in silicon nanoindentation", *Phys. Rev. Lett.* **84**, 1260–1263 (2000).

[SV00] M. R. Sorensen and A. Voter, *J. Chem. Phys.* **112**, 9599–9606 (2000).

[TOP96] E. B. Tadmor, M. Ortiz and R. Phillips, "Quasicontinuum analysis of defects in crystals," *Phil. Mag. A* **73**, 1529–1563 (1996).

[TK98] S. Tiwari and A. Klar, an adaptive domain decomposition procedure for Boltzmann and Euler equations, *J. Comput. Appl. Math.* **90**, 223–237 (1998).

[VE03] E. Vanden-Eijnden, "Numerical techniques for multiscale dynamical systems with stochastic effects", *Comm. Math. Sci.*, **1**, 385–391 (2003).

[Vot97] A. Voter, "Hyperdynamics: Accelerated molecular dynamics of infrequent events", *Phys. Rev. Lett.* **78**, 3908–3911 (1997).

[WL03] G.J. Wagner and W.K. Liu "Coupling of atomistic and continuum simulations using a Bridging Scale Decomposition", preprint.

[WL76] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions", *J. Mol. Biol.* **103**, 227–249 (1976).

[Xu94] K. Xu and K.H. Prendergast, "Numerical Navier-Stokes solutions from gas kinetic theory," *J. Comput. Phys.* **114**, 9–17 (1994).

[ZLY99] Y. Zhang, T.-S. Lee and W. Yang, "A pseudobond approach to combining quantum mechanical and molecular mechanical methods", *J. Chemical Phys.* **110**, 46–54 (1999).

Homogenization Method for Transport of DNA Particles in Heterogeneous Arrays

Assyr Abdulle[1] and Sabine Attinger[2]

[1] CoLab, Computational Laboratory, ETHZ ** abdulle@inf.ethz.ch
[2] CoLab, Computational Laboratory, ETHZ sabine.attinger@inf.ethz.ch

In this paper we study the large scale transport of the DNA particles through a heterogeneous micro array in the framework of homogenization theory. We derive the macro scale particle transport equation and show that for transport of particles in a divergence free electric field as proposed by Duke and Austin [Du98] and Ertas [Er98] separation according to particle mass or size cannot been achieved. Our results explain the experimental findings of Duke and Austin [Du98] and Ertas [Er98] and thus close the gap between theory and experiment.

1 Introduction

Separation of large biomolecules such as DNA and proteins is of high interest for biological research and biomedical application. A technique currently used is the separation through gel electrophoresis in which different molecules of different size migrate at different speed through the gel. A new approach has been suggested by Duke and Austin [Du98] and Ertas [Er98] only recently. The authors proposed to sort molecules by injecting them in an electric field modified by an array of spatially asymmetric obstacles. In their articles, the authors claim that due to the asymmetry of the array only, particles are deflected away from the mean direction of the electric field. Since this effect depends on diffusive motion, the macromolecules should be sorted in transverse direction according to their size. The authors consider their approach as a simple realization of a "thermal ratchet" [Ma93, Pr94]. These later devices separate diffusive particles by switching on and off an asymmetric potential.

However, the experimental setup in [Hu02] showed that particles much smaller than the barrier gap follow individual flow lines through narrow gaps and are poorly fractionated. In contrast, particles comparable to the gap size can be fractionated with high resolution. Austin et al. [Au02] give qualitative arguments that in a divergence free flow field no separation occurs. Only if there are additional forces acting on the particles as it is the case for the larger particles that are strongly interacting with the

** Now at: Department of Mathematics, University of Basel

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

impervious obstacles fractionation might be achieved. A closed theory is still missing in order to explain this behaviour and we aim at closing this gap.

Since the conductivity of the sorting array can be modeled as being periodic, with period given by the gap between two obstacles, the particle transport equation is a partial differential equation of two-scale nature, with a microscopic length and a macroscopic length. A natural framework to treat well separated multi-scale equations is the homogenization theory which describes the replacement of a two-scale equation by a homogenized equation where small scales have been averaged out.

In this paper we study the transport of the DNA particles in the framework of homogenization theory. We will show that with a divergence free flow field, the macroscopic drift of the homogenized equation is independent of the diffusion of the macromolecules and proper separation based on different diffusion constant seems unlikely to happen. We believe that this gives the explanation why the small macromolecules can not be well separated by a microarray sieve. In a future work, we will study the case of a non divergence free electrical flow field, which would allow to find a drift depending on the diffusion coefficients of the macro molecules.

2 The fine scale particle transport

We first constitute the conservation law for electrically charged particles that move in an electrical field E, the gradient of the electrical potential ϕ and $E = \nabla\phi$. The electrical potential is determined by

$$\nabla \cdot (k\nabla\phi) = 0 \tag{1}$$

where k is the electrical conductivity. The flux of DNA particles J_E due to the electrical field E is given by $J_E = \hat{c}\mu E$ where $\hat{c}$ denotes the DNA particle concentration and μ is the mobility. The mobility μ is related to the electrical conductivity by $\mu = \rho k$ with ρ the charge density of the electrical array which we assume to be constant. In contrast to standard electrophoresis where separation is achieved due to particle size dependent mobility values, here the mobility is assumed to depend only on the geometry of the micro array. To obtain the total particle flux, a diffusive flux is added

$$J = \hat{c}\mu E - D\nabla\hat{c}, \tag{2}$$

so that the mass conservation law for the particle concentration reads

$$\frac{\partial\hat{c}}{\partial\hat{t}} = -\nabla(\hat{c}\mu E) + \nabla(D\nabla)\hat{c}. \tag{3}$$

Assuming a divergence free electrical field and an isotropic diffusion tensor D and defining a velocity field $\hat{u}$ as $\hat{u} \equiv \mu E$, we finally obtain a conservation law in the form of an advection diffusion equation

$$\frac{\partial\hat{c}}{\partial\hat{t}} = -\hat{u}\nabla\hat{c} + D\nabla^2\hat{c}. \tag{4}$$

There are two typical length scales involved in our problem. A small length scale l which represents the distance between neighbouring obstacles of the device and a large length scale L at which the transport behavior is observed. We therefore introduce two different spatial variables: $\hat{x}$ for large scale and $\hat{y}$ for small scale variations. A non-dimensional representation of equation (4) follows by scaling the $\hat{x}, \hat{y}, \hat{t}$ variables according to $x = \hat{x}/L, y = \hat{y}/l, t = \hat{t}D/L^2$. c denotes the non-dimensional concentration. By scaling the velocity field $\hat{u}$ as $\hat{u}/U := u$, where U is the norm of a characteristic macroscopic velocity, we obtain

$$\frac{\partial c}{\partial t} = -P_L \, u\nabla c + \nabla^2 c, \tag{5}$$

where the dimensionless parameter P_L (the Péclet number), defined as $P_L = UL/D$, sets the diffusion coefficient into relation with transport velocity and the large length scale of the problem. The capital subscript L stands for the scaling with respect to the macro scale length L. The ratio between the characteristic macroscopic and microscopic length scales will be denoted by $\varepsilon = l/L$. Thus, we have the relation $y = x/\varepsilon$.

We will use in the sequel the relation

$$P_L = \frac{UL}{D} = \frac{Ul}{D}\frac{L}{l} = P_l \frac{1}{\varepsilon}, \tag{6}$$

depending wether we rescale with respect to the macro length L or the micro length l. Moreover, we will add a superscript ε to the concentration c and the velocity field u to emphasize the dependence on ε. Thus the nondimensional multiscale equation (5) reads

$$\frac{\partial c^\varepsilon}{\partial t} = -P_L \, u^\varepsilon \nabla c^\varepsilon + \nabla^2 c^\varepsilon, \tag{7}$$

3 Homogenization of particle transport

In this section we will derive the homogenized solution of equation (5). Homogenization theory is able to extract the macroscopic system behaviour from a system that is governed by microscopic heterogeneities. For our problem, the heterogeneities at the microscopic level are the obstacles of the periodic asymmetric array, the macroscopic transport behaviour originates from the transport equation (5) but describes the transport of the particles on a macro length scale where the array appears homogeneous. We assume that the concentration c^ε and the velocity field u^ε depend on both the macro and micro scale x, y respectively and that there are 1-periodic (for simplicity) with respect to the y variable. By $Y = (0,1)^2$ we will denote the y domain. Note that by setting $y = x/\varepsilon$ the above functions are ε periodic with respect to the second variable.

We first split the velocity field u^ε into a large scale part u_0 and a fluctuating part $\tilde{u}$,

$$u^\varepsilon(x) = u(x,y) = u_0(x) + \tilde{u}(x,y), \tag{8}$$

where $u_0(x) = \int_Y u(x,y)dy$ and $y = x/\varepsilon$. A consequence is that the fluctuating part satisfies $\int_Y \tilde{u}(x,y)dy = 0$. Since the concentration depends on both scales $x, y = x/\varepsilon$, application of the chain rule yields

$$\nabla = \nabla_x + 1/\varepsilon \nabla_y. \tag{9}$$

We obtain then for equation (5)

$$\frac{\partial c^\varepsilon}{\partial t} = -P_L u_0 \nabla_x c^\varepsilon + \nabla_x^2 c^\varepsilon - \frac{1}{\varepsilon} P_l \tilde{u} \nabla_x c^\varepsilon - \frac{1}{\varepsilon^2} P_l u^\varepsilon \nabla_y c^\varepsilon + 2\frac{1}{\varepsilon} \nabla_x \nabla_y c^\varepsilon + \frac{1}{\varepsilon^2} \nabla_y^2 c^\varepsilon, \tag{10}$$

where the first two terms of the right-hand side represent macroscopic drift and diffusive flux terms scaled with the macroscopic length L. All other terms describe microscopic processes and are scaled with the microscopic length l.

Remark 1. The above scaling motivated by physical considerations can be also justified by perturbation theory arguments, which demonstrates that the scaling described above is the only one to obtain the correct homogenized equation [De00].

Next, we assume a two scale expansion for the concentration

$$c^\varepsilon(t,x) = c(t,x,y) = c_0(t,x) + \varepsilon c_1(t,x,y) + \varepsilon^2 c_2(t,x,y) + \ldots, \tag{11}$$

where c_0 is the homogenized solution we are looking for (and thus does not depend on the micro scale $y = x/\varepsilon$) and where we assume that c_j are periodic in the y variable. The above ansatz is a formal expansion, the interest in it stems from the fact that it allows to derive formally an equation for the homogenized solution (the coarse description). The procedure is now to introduce (11) into (10) and to collect the terms with the same power in ε. We then obtain the following set of equations

$$\varepsilon^{-1} : -\nabla_y^2 c_1 + P_l u^\varepsilon \nabla_y c_1 = -P_l \tilde{u} \nabla_x c_0, \tag{12}$$

$$\varepsilon^0 : -\nabla_y^2 c_2 + P_l u^\varepsilon \nabla_y c_2 = -P_L u_0 \nabla_x c_0 + \nabla_x^2 c_0 - P_l \tilde{u} \nabla_x c_1 \tag{13}$$

$$+2\nabla_x \nabla_y c_1 - \frac{\partial c_0}{\partial t}. \tag{14}$$

The equation for ε^{-1} determines c_1 in terms of c_0 in the following way: let $\chi_i(y),\ i = 1,2$ be the solution of the so called cell problem

$$\nabla_y^2 \chi_i - P_l u^\varepsilon \nabla_y \chi_i = P_l \tilde{u}_i \tag{15}$$

in the domain Y with periodic boundary conditions then by linearity and variable separation of equation (15), the solution of (12) is given by

$$c_1 = \chi \cdot \nabla_x c_0, \tag{16}$$

where $\chi = (\chi_1, \chi_2)$. The equation (13) can be considered as an equation for the unknown c_2 in the domain Y with periodic boundary conditions. A necessary condition for this equation to have a solution is that the average of the remainder term

(which do not involve c_2) vanishes (see [Ev98] for details). This condition together with inserting the expression (16) for c_1 gives the homogenized equation

$$\frac{\partial c_0}{\partial t} = -P_L u_0 \nabla_x c_0 + \nabla_x \cdot I_0 \nabla_x c_0, \tag{17}$$

where I_0 is an effective diffusion coefficient and defined by

$$I_0 = I + \int_Y P_l \tilde{u} \otimes \chi dy \tag{18}$$

and where I is the identity matrix and

$$\tilde{u} \otimes \chi = \begin{pmatrix} \tilde{u}_1\chi_1 & \tilde{u}_1\chi_2 \\ \tilde{u}_2\chi_1 & \tilde{u}_2\chi_2 \end{pmatrix}. \tag{19}$$

To solve the homogenized equation one has to solve first the cell problems (15) and then the macro equation (17). If one goes back to a dimensional characterization, and substitute $\hat{x} = xL, \hat{t} = tL^2/D$ for x, t and $\hat{u} = uU$ for u we obtain for the homogenized equation

$$\frac{\partial \hat{c}_0}{\partial \hat{t}} = -\hat{u_0} \nabla_{\hat{x}} \hat{c}_0 + \nabla_{\hat{x}} \cdot \hat{I}_0 \nabla_{\hat{x}} \hat{c}_0, \tag{20}$$

where $\hat{I}_0$ now is defined by

$$\hat{I}_0 = D + \int_Y \hat{\tilde{u}} \otimes \chi dy. \tag{21}$$

Thus we have transferred the heterogeneous fine scale model (5) into a homogeneous large scale model (20) which describes the macroscopic behaviour of the particles through the micro array (microscopically heterogeneous due to the microscopic obstacles). We see in equation (20) the large scale drift $\hat{u}_0$ does not depend on the diffusion constant or the molecular weight of the particles. Thus particles with different weight, i.e. different diffusion coefficients, will move with the same drift. The heterogeneous micro array does have an impact but only in form of an effective diffusivity.

We conclude that for particle transport in heterogeneous divergence free flow fields, no diffusion dependent deflection of particles from the mean flow direction exists and thus no particle separation can occur. It explains the experimental findings presented in [Hu02].

3.1 Solution of the cell problem in perturbation theory

There is no simple closed solution for the cell problem (15) which we need to solve before evaluating the macrodiffusion tensor (18). However, for the micro array with system parameters chosen according to Duke [Du98] and Ertas [Er98] the Peclet number P_l is large and the variance of the electric conductivity fluctuations is small.

Thus, a perturbation theory expansion with respect to small conductivity fluctuations and small inverse peclet numbers will yield a reliable approximate solution for the cell problem. In lowest order approximation the cell problems reduces to

$$-P_l u_0 \nabla_y \chi_i = P_l \tilde{u}_i \tag{22}$$

Following the lines of [Lu02], we explicitly compute the entries of the macrodiffusion tensor by evaluating the integral expressions in

$$\hat{I}_{0\eta_i,\eta_j} = D + u_0 \int_0^\infty R_{u_{\eta_i} u_{\eta_j}}(\eta') d\eta' \tag{23}$$

in a curved coordinate system η with coordinate axes aligned in mean flow direction η_1 and transverse to it, η_2. $R_{u_{\eta_i} u_{\eta_j}}$ is the dimensionless velocity correlation function. It is determined by the statistics of the electric conductivity via Ohm's law (27). In a uniform flow field, the curved coordinate system reduces to the Cartesian coordinate system and the macrodiffusion coefficients (23) reduce to the formulae known since the pioneering work of Gelhar and Axness [Ge83] and Dagan [Da89].

The macrodiffusivity (23) depends on the angle between the mean drift and the principal axes of the heterogeneous structure of the medium. The macrodiffusion tensor becomes diagonal in a coordinate system along the streamlines, but its entries are not isotropic. The component parallel to the mean flow direction increases through interaction with the heterogeneities whereas the component transverse to it is approximately given by the microscopic diffusion coefficient D.

The explicit evaluation of the macrodiffusion coefficient yields

$$\hat{I}_{0\eta_1,\eta_1} = D + u_0 \int_0^\infty R_{u_{\eta_1} u_{\eta_1}}(\eta_1') d\eta_1' = D + u_0 \sigma_k^2 l_{\eta_1} = D + 0.00035 \tag{24}$$

$$\hat{I}_{0\eta_2,\eta_2} = D \tag{25}$$

where l_{η_1} is the integral scale of the micro array in mean flow direction [Lu02] and where σ_k is the variance of the electric conductivity (27).

4 Numerical Simulations

In this section, we confirm our theoretical results by numerical computations. The direct solution of equation (5) is costly due to dependence of the problem on the small length scale. For standard methods, the mesh has to be sufficiently fine to capture the heterogeneities of the micro array, leading to a large number of unknowns. Furthermore, the velocity field u^ε (see (8)), which is computed from the electrostatic potential ϕ^ε (see equation (27) below) via an elliptic equation, also exhibits multiple scales. Here also a fine mesh is needed to gain full resolution. Replacing the heterogeneous micro array by the homogenized one and computing the particle transport on it makes it possible to reduce the computational resolution by choosing a coarser mesh.

However, to test the reliability of homogenization in our case we will determine in a first step the fine scale solution and compare it afterwards with the homogenized solution.

4.1 Numerical Parametrization

We recall that the transport equation for the concentration is given by (see (4))

$$\frac{\partial c^\varepsilon}{\partial t} = -u^\varepsilon \nabla c^\varepsilon + D\nabla^2 c^\varepsilon, \tag{26}$$

where the velocity $u^\varepsilon = \rho\, k^\varepsilon \nabla \phi^\varepsilon$ is the solution of a Darcy type equation for the electrical potential ϕ in $\Omega = (0,1)^2$

$$\nabla \cdot (k^\varepsilon \nabla \phi^\varepsilon) = 0 \text{ in } \Omega \tag{27}$$

$$\phi^\varepsilon(0, x_2) = \phi^\varepsilon(x_1, 1) = 1, \tag{28}$$

$$\phi^\varepsilon(x_1, 0) = \phi^\varepsilon(1, x_2) = 0. \tag{29}$$

As usual we add a superscript ε to the concentration c, the velocity field u, the potential ϕ and the conductivity k to indicate the dependence of these quantities on the micro array parameters (see Figure 1 below).

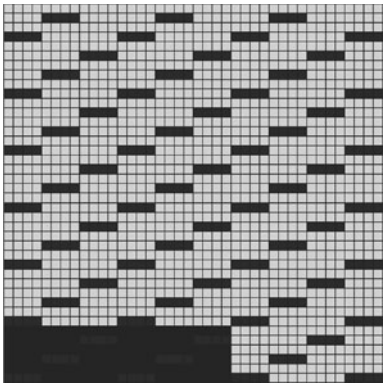


Fig. 1. Sample of the microarray. (See Plate 2 on Page 269)

We discretize Ω in 400 equidistant steps in each spatial direction x_1, x_2. A sample of the asymmetric periodic array is sketched below, where the periodic unit cell is indicated by the shaded area. It is a rectangle of size $2\varepsilon \times \varepsilon$. Thus for our example, $\varepsilon \simeq 0.0150$. We chose $k^\varepsilon = 0.1$ inside and $k^\varepsilon = 0$ outside the obstacles and set $\rho = 1$.

In order to minimize the influence of artificial boundaries conditions (which we take as zero Dirichlet), we perform the transport simulation on a subdomain $\Omega_s = [0.375, 0.625] \times [0.375, 0.625]$ with 100 steps in each spatial direction x_1, x_2.

4.2 Numerical Methods

The elliptic equation is solved with a standard finite element method in order to compute the fine scale reference solution for the velocity field $u^\varepsilon = k^\varepsilon \nabla \phi^\varepsilon$. The advection diffusion equation (26) is solved by the method of lines. In this method, the spatial variables are discretized by finite difference, and the resulting equations are then a system of ordinary differential equations (ODEs). It is known that standard explicit solvers for such system arising from reaction-diffusion equations exhibit time step restrictions due to stability problems. A possibility is then to use an implicit solver with the drawback of having to solve large linear systems [Ha96]. In some situations, as for diffusion dominated problems, a class of methods called Chebyshev methods can be very efficient. These methods are explicit and at the same time possess large stability domains along the negative real axis. Recently, a Chebyshev method of high order, ROCK4 has been proposed in [Ab02]. This method is used in the numerical experiments.

For homogenizing the elliptic equation (27) and thus for determining the large scale drift, we used the finite element heterogeneous multiscale method (FE-HMM) described in [Ab03], based on a methodology introduced in [We02]. We used it as a procedure to calculate the homogenized conductivity values without deriving analytically the homogenized equations. Heterogeneous multiscale method (HMM) introduced in [We02] is a general framework for the numerical computation of solutions to problems with multiple scales. For homogenization problems, its main objective is to avoid the precomputation of fine scale shape functions as well as of the coefficients of the homogenized equation. It can in this context be seen as a numerical homogenization procedure where in addition, it is possible to recover fine scale information.

5 Results

Fine Scale Simulation
In the fine scale simulation, we compute the fine scale particle transport (26) through a heterogeneous array (k^ε given above) and compare it with the solution for particle transport through a fictitious isotropic homogeneous array ($k = 0.1$).

We see in Figures 2 and 3 that the flow lines are distracted from the 45^o diagonal, i.e., the effective conductivity tensor is anisotropic. It implies that an homogeneous array can not capture this effect. We also see that for two different microscopic diffusion coefficients the large scale center of mass of the particle cloud does not depend on them. This implies that there is no separation of particles based on diffusion coefficient in this way.

Large Scale Simulation
In this simulation, we compare the heterogeneous transport simulation with the homogenized transport simulation with the effective conductivity tensor k^0 derived by

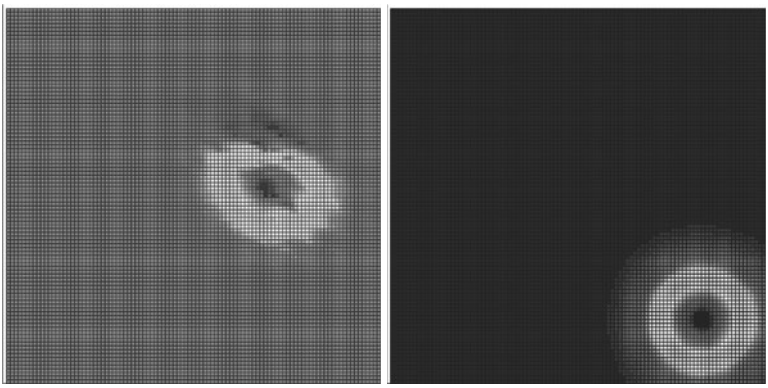


Fig. 2. Transport simulation in a homogeneous array versus heterogeneous array, $k = 0.1, D = 0.0005$. (See Plate 3 on Page 270)

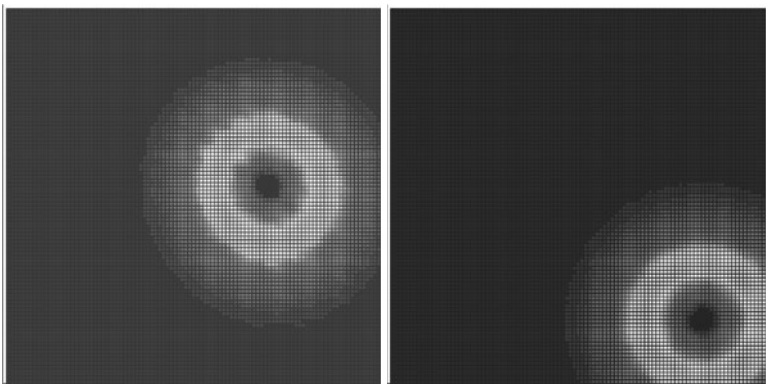


Fig. 3. Transport simulation in a homogeneous array versus heterogeneous array, $k = 0.1, D = 0.001$. (See Plate 4 on Page 270)

homogenization of the elliptic equation for the potential (see [Be78] for the derivation of the homogenized equations). The computation of k^0 is done with the FE HMM implementation given in [Ab03]. We obtain for the effective conductivity tensor

$$\begin{pmatrix} 0.08074 & 0 \\ 0 & 0.04588 \end{pmatrix}$$

As observed in the above numerical computations, the micro array induces an anisotropic effective conductivity tensor. The anisotropy ratio is given by 0.5682.

We gave the formula for the effective diffusion coefficients in a coordinate system aligned parallel to the mean flow direction in (24). The transverse effective diffusion

coefficient is $D = 0.0001$ or $D = 0.005$ respectively. Its rotation in Cartesian coordinates yields the diffusion matrices

$$\begin{pmatrix} 0.00109(0.00059) & 0.00015 \\ 0.00015 & 0.00127(0.00077) \end{pmatrix}$$

The values without brackets belong to the effective diffusion matrix with the larger microscopic diffusion $D = 0.005$ whereas the values in brackets belong to the smaller microscopic diffusion $D = 0.0001$.

The anisotropic homogeneous array captures now the distraction from the 45^0 diagonal. We see in Figures 4 and 5 that flow lines of the homogenized array are a slightly more distracted than those of the heterogeneous array. The reason is that the particles first need to sample some correlation lengths of the heterogeneous array

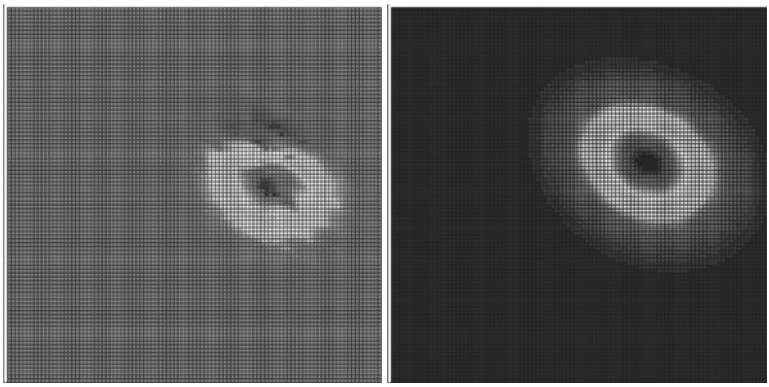


Fig. 4. Transport simulation for the homogenized array versus heterogeneous array, $D = 0.0005$. (See Plate 5 on Page 270)

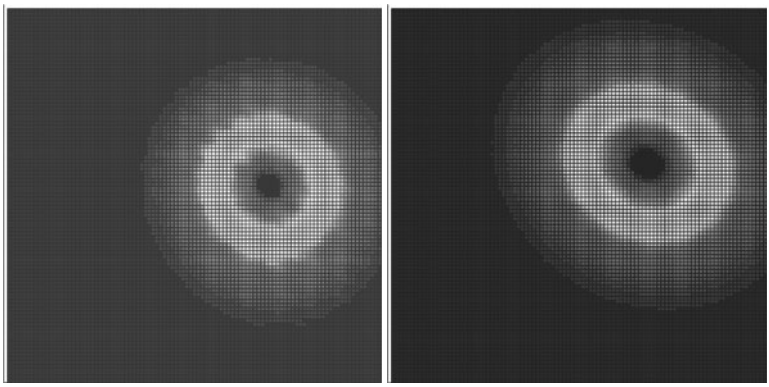


Fig. 5. Transport simulation for the homogenized array versus heterogeneous array, $D = 0.001$. (See Plate 6 on Page 271)

in order to approach their asymptotic distraction angle. Therefore, the path of the moving particle cloud in the heterogeneous array is shifted against the path in the equivalent homogeneous array. However, in the asymptotic regime the distraction angle is the same and the shift is a constant indicating that homogenization yields reliable results.

Moreover, we see again that the center of mass of the particle cloud does not depend on the diffusion coefficient and that there is no separation of particles based on diffusion coefficient in this way.

References

[Ab02] A. Abdulle, *Fourth order Chebyshev methods with recurrence relation*, SISC, Vol.23, N0.6, pp. 2041–2054, 2002.

[Ab03] A. Abdulle and C. Schwab, *Heterogeneous multi-scale FEM for diffusion problem on rough surfaces*, submitted to SIAM MMS.

[Be78] A. Bensoussan, J.-L. Lions and G. Papanicolaou, *Asymptotic Analysis for Periodic Structures*, North Holland, Amsterdam, 1978.

[Du98] T.A. Duke and R.H. Austin, *Microfabricated Sieve for the Continuous Sorting of Macromolecules* Phys. rev. Lett. Vol. 89, No. 7, 1998.

[We02] W. E and B. Engquist, *The Heterogeneous Multi-Scale Methods*, Comm. Math. Sci., Vol. 1, No. 1, pp. 87-132, 2003.

[Ev98] L.C. Evans, *Partial Differential Equations*, AMS, Providence, Rhode Island, 1998.

[Er98] D. Ertas, *Lateral Separation of Macromolecules and Polyelectrolytes in Microlithographic Arrays* Phys. Rev. Lett. Vol. 80, No. 7, 1998.

[Ha96] E. Hairer, G. Wanner, *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems.* Springer Verlag Series in Comput. Math., Vol. 8, Berlin, 1996.

[Hu02] L.R. Huang, P. Silberzan, J.O. Tegenfeldt, E.C. Cox, J.C. Sturm, R.H. Austin and H. Craighead, *Role of Molecular Size in Ratchet Fractionation*, Phys. rev. Lett. Vol. 89, No. 17, 2002.

[Au02] R.H. Austin, N.Darnton, R. Huang, J. Sturm, O. Bakajin, T. Duke *Ratchets: the problems with boundary conditions in insulating fluids*, Appl. Phys. A 75, 279284, 2002.

[Ma93] M. Magnasco, *Forced Thermal Ratchets*, Phys. Rev. Lett. 71, 1477, 1993.

[Pr94] J. Prost, J.F. Chaudwin, L. Peliti, and A. Ajdari, *Asymmetric Pumping of Particles*, Phys. Rev. Lett. 72, 1766, 1994.

[Ge83] Gelhar, L.W., and C.L. Axness, *Three-Dimensional stochastic analysis of macrodispersion in aquifers*, Water Resour. Res., Vol 19, 161-180, 1983.

[Da89] G. Dagan, *Flow andTransport in Porous Formations*, Springer-Verlag, New York, 1989.

[De00] M. Dentz, H. Kinzelbach, S. Attinger and W. Kinzelbach, *Temporal behaviour of a solute cloud in a heterogeneous porous medium: Point like injection*, Water Resour. Res., Vol. 36, No. 12, 3591-3604, 2000

[Lu02] I. Lunati, S. Attinger and W. Kinzelbach, *Large Scale Dispersivities for Transport in arbitrary Nonuniform Flow Fields*, Water Resour. Res., Vol 38, No.10, 1187 , 2002.

Metastability, conformation dynamics, and transition pathways in complex systems

Weinan E[1] and Eric Vanden-Eijnden[2]

[1] Department of Mathematics and PACM, Princeton University, Princeton, NJ 08544
weinan@princeton.edu
[2] Courant Institute, New York University, New York, NY 10012 eve2@cims.nyu.edu

Summary. We present a systematic introduction to the basic concepts and techniques for determining transition pathways and transition rates in systems with multiple metastable states. After discussing the classical transition state theory and its limitations, we derive a new set of equations for the optimal dividing surfaces. We then discuss transition path sampling, which is the most general technique currently available for determining transition regions and rates. This is followed by a discussion on minimal energy path for systems with smooth energy landscapes. For systems with rough energy landscapes, our presentation is centered around the notion of reaction coordinates. We discuss the two related notions of free energies associated with a reaction coordinate, and show that at least in the high friction limit, there does exist an optimal reaction coordinate that gives asymptotically the correct prediction for the transition rates. Variational principles associated with the optimal reaction coordinates are exploited under the assumption that the transition paths are restricted to tubes, and this provides a theoretical justification for the finite temperature string method. Blue moon sampling techniques, metadynamics and a new form of accelerated dynamics are also discussed.

1 Introduction

The evolution of complex systems often involves widely separated time-scales. Well-known examples include nucleation events during phase transition, conformational changes of molecules, and chemical reactions. The appearance of long time-scales is related to the existence of energy barriers or entropic bottlenecks which partition the system into metastable basins. A dynamical trajectory may spend a very long time in one basin before finding its way to another. The separation of time-scales is typically so pronounced that it is difficult to observe a single transition event, let alone gather enough statistical information about the transitions, by using conventional dynamical simulations. Several techniques have been introduced to tackle the numerical difficulty of determining the transition pathways and transition rates in systems of this kind. The purpose of the present paper is to review some (not all) of these techniques and the underlying theoretical framework. In doing so, we find it sometimes more convenient to take a point of view that is slightly different from that of the original papers. In addition, we will make an attempt to put frequently used notions such as

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

reaction coordinates, optimal reaction coordinates and optimal dividing surfaces on a solid footing.

We shall mostly focus on methods which assume the knowledge of the initial and final states for the transition that we are interested in. The transition may proceed through additional metastable states that are not identified beforehand. We will understand metastability in the following way. Suppose that the system admits a unique equilibrium distribution such as the Gibbs distribution

$$\rho(x) = Z^{-1} e^{-\beta V(x)}, \qquad Z = \int_{\mathbb{R}^n} e^{-\beta V(x)} dx \tag{1}$$

where $x \in \mathbb{R}^n$ is the configuration space, $V(x)$ is some given potential, and $\beta = 1/k_B T$ is the inverse temperature. Then the regions $\{B_j\}_{j=1}^J$ will be metastable if

$$\sum_{j=1}^{J} N_j \approx 1 \qquad \text{where } N_j = \int_{B_j} \rho(x) dx, \tag{2}$$

i.e. with very high probability, the system at equilibrium is found in one of the sets B_j. Assuming ergodicity, N_j is also the fraction of time that the system spends in B_j. In addition to (2) one must also require that the regions B_j be well separated so that the transitions between these regions are quasi-Markov events. We shall come back to this requirement later. Note that the Gibbs distribution in (1) is consistent with the Langevin dynamics

$$\begin{cases} \dot{x} = v \\ M\dot{v} = -\nabla V(x) - \gamma v + \sqrt{2\gamma\beta^{-1}}\,\eta \end{cases} \tag{3}$$

where $\eta(t)$ is a white-noise, γ is the friction coefficient, and M is the diagonal mass matrix. Two limiting cases of (3) are the overdamped (high friction) dynamics obtained as $\gamma \to \infty$

$$\gamma\dot{x} = -\nabla V(x) + \sqrt{2\gamma\beta^{-1}}\,\eta \tag{4}$$

and the Hamiltonian dynamics which arises when $\gamma = 0$

$$M\ddot{x} = -\nabla V(x). \tag{5}$$

2 Transition state theory and the Bennett-Chandler procedure

Transition state theory is the oldest attempt to describe metastability in ergodic systems [Wig38, Eyr35]. Consider a system governed by the Hamiltonian dynamics (5), and let us partition the configuration space into two sets, A and its complement A^c. Define the mean residence time in A and A^c as

$$t_A = \lim_{T\to\infty} \frac{2}{N_T} \int_0^T \chi_A(x(t)) dt, \tag{6}$$

and similarly for t_{A^c}. Here $x(t)$ is the instantaneous position of a generic trajectory, N_T is the number of times this trajectory crosses the boundary between A and A^c before time T, and χ_A denotes the indicator function of A, i.e. $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ otherwise. (6) can be rewritten as

$$t_A = N_A/\nu^{\text{TST}}, \qquad t_{A^c} = N_{A^c}/\nu^{\text{TST}} \tag{7}$$

where

$$N_A = \lim_{T\to\infty} \frac{1}{T} \int_0^T \chi_A(x(t))dt = \int_A \rho(x)dx, \tag{8}$$

is the proportion of time that the trajectory $x(t)$ spends in A, i.e. the equilibrium population density in A assuming ergodicity, $N_{A^c} = 1 - N_A$, and

$$\nu^{\text{TST}} = \lim_{T\to\infty} \frac{N_T}{2T}, \tag{9}$$

is the (half) mean frequency of crossing the boundary ∂A. This mean frequency can be estimated upon noting that $|\dot{\chi}_A(x(t))|$ is a sum of delta functions concentrated at the times when $x(t)$ crosses ∂A. Therefore

$$\begin{aligned} \nu^{\text{TST}} &= \lim_{T\to\infty} \frac{1}{2T} \int_0^T |\dot{\chi}_A(x(t))|dt \\ &= \frac{1}{2} \int_{\mathbb{R}^n} \int_{\partial A} |\hat{n} \cdot v| \rho(x,v) d\sigma(x) dv \end{aligned} \tag{10}$$

Here $\rho(x,v)$ is the joint equilibrium density of (x,v) and $d\sigma(x)$ is the surface element on the boundary ∂A. Assuming that the number of degrees of freedom is large, $\rho(x,v)$ can be approximated by the Gibbs distribution $Z_H^{-1} e^{-\beta(|v|^2/2+V(x))}$, and (10) reduces to

$$\nu^{\text{TST}} = \sqrt{\frac{2}{\pi\beta}} Z^{-1} \int_{\partial A} e^{-\beta V(x)} d\sigma(x). \tag{11}$$

Transition state theory is exact for the computation of the mean residence time. But the mean residence time may not be very relevant as far as description of transition events is concerned. Indeed, since a trajectory that has just crossed the boundary ∂A may have a high probability of re-crossing it right away, the successive transition times between A and A_c are correlated in general. Therefore, the symbolic dynamics on these two sets is usually poorly approximated by a two-state Markov chain with rates

$$k_{AA^c} = t_A^{-1} = \nu^{\text{TST}}/N_A, \qquad k_{A^cA} = t_{A^c}^{-1} = \nu^{\text{TST}}/N_{A^c} \tag{12}$$

Quasi-Markov transitions will only arise between sets that are sufficiently separated, i.e. there exists a buffer region between them that has a large volume but negligible probability. How to correct transition state theory to handle this situation was explained by Bennett [Ben77] and Chandler [Chan78] (see also [Yam60]). These authors base their considerations on response functions. Here we take a slightly different viewpoint. Consider a system with two metastable sets, B_1 and B_2, satisfying (2)

with $J = 2$, and define A such that ∂A belongs to the buffer region $(B_1 \cup B_2)^c$ (i.e. $B_1 \subset A$ and $B_2 \subset A^c$). Then the rates between B_1 and B_2 can be expressed as

$$k_{12} = \nu/N_1, \qquad k_{21} = \nu/N_2, \tag{13}$$

where $N_1 \approx N_A$ and $N_2 \approx N_{A^c}$ are the population densities in B_1 and B_2, respectively, and ν is the corrected frequency

$$\nu = \frac{1}{2}\int_{\mathbb{R}^n}\int_{\partial A} |\hat{n}\cdot v|\rho(x,v)\chi(x,v)d\sigma(x)dv. \tag{14}$$

The factor $\chi(x,v)$ is included so that only trajectories passing through (x,v) which correspond to transition events, i.e. they go from ∂B_1 to ∂B_2, or ∂B_2 to ∂B_1, are accounted for. Thus, if $N_R(x,v)$ is the number of crossings of ∂A between the times the trajectory exit B_1 and enters B_2 or conversely, one has

$$\chi(x,v) = \tfrac{1}{2}\left(1-(-1)^{N_R(x,v)}\right)/N_R(x,v). \tag{15}$$

The first factor $\frac{1}{2}(1-(-1)^{N_R})$ is 1 if N_R is odd (i.e. if the trajectory connects B_1 to B_2 or B_2 to B_1) and zero otherwise; the second factor N_R^{-1} accounts for over-counting since N_R starting points on ∂A are contained in the same transition trajectory. The ratio ν/ν^{TST}, which is necessarily smaller or equal to one, is called the transmission coefficient κ of the surface ∂A.

The Bennett-Chandler approach in principle offers a way of identifying the transition paths and computing the mean transition times in a two-step procedure. First choose a dividing surface ∂A and compute the expectation in (10) or (11) using e.g. umbrella or blue-moon sampling techniques and thermodynamic integration (see section 8). Then evaluate the transmission coefficient of ∂A using

$$\kappa = \tfrac{1}{2}\left\langle\left(1-(-1)^{N_R(x,v)}\right)/N_R(x,v)\right\rangle_{\partial A} \tag{16}$$

where $\langle\cdot\rangle_{\partial A}$ denotes expectation with respect to the density proportional to $|\hat{n}\cdot v|\rho(x,v)$ after proper normalization. Evaluating (16) amounts to initiating trajectories from ∂A based on this density, running them both forward and backward in time, observing if they correspond to transition trajectories and counting the number of times they re-cross ∂A. Unfortunately, this procedure suffers from the following practical difficulties.

1. ∂A might be a poor dividing surface in the sense that very few trajectories initiated on ∂A give rise to switching and, hence, are statistically significant in the Bennett-Chandler procedure. A good dividing surface should have the property that trajectories initiated on it have about half probability to reach either B_1 or B_2.

2. Even if ∂A is a good dividing surface, it will be difficult to compute accurately the transmission coefficient κ if the transition trajectories tend to re-cross many times ∂A before reaching either B_1 or B_2. The main reason is that the initial conditions

generated on ∂A by sampling are correlated when the number of re-crossings is large. A set of N transition trajectories that on the average re-cross N_R times the surface ∂A contains only N/N_R statistically significant transition trajectories [Str01]. In situations when N_R is large, the transition is said to be diffusive (as opposed to being ballistic when recrossing is rare). This may mean that the transition state is in reality a rather wide region.

3 Optimizing TST

Since TST necessarily overestimate the rate, it is natural to look for the dividing surface with minimum TST rate. Using (7) and (11) this is the dividing surface that minimizes

$$I_0 = \int_{\partial A} e^{-\beta V} d\sigma(x). \tag{17}$$

It is shown in Appendix A that the minimizer of (17) satisfies

$$0 = -\nabla V \cdot \hat{n} + \beta^{-1}\kappa, \tag{18}$$

where $\hat{n}(x)$ is the unit normal to ∂A at point x and $\kappa = \nabla \cdot \hat{n}$ is the local mean curvature of this surface.

In general, optimizing ∂A by solving (18) will be too formidable a task to be practical. More realistically, one can minimize the functional (17) over restricted classes of surfaces. The simplest choice is to assume that ∂A is planar, in which case it is specified by the equation

$$0 = \hat{n} \cdot x - b \tag{19}$$

where b is a scalar and $\hat{n}$ is the unit normal to the plane. Using the planar ansatz in (17) gives

$$I_0 = \int_{\mathbb{R}^n} e^{-\beta V} \delta(\hat{n} \cdot x - b) dx. \tag{20}$$

As shown in Appendix A, the minimizers of (20) satisfy

$$\begin{cases} 0 = \displaystyle\int_{\mathbb{R}^n} \hat{n} \cdot \nabla V e^{-\beta V} x^{\perp}\, \delta(\hat{n} \cdot x - b) dx \\ 0 = -\displaystyle\int_{\mathbb{R}^n} \hat{n} \cdot \nabla V e^{-\beta V} \delta(\hat{n} \cdot x - b) dx, \end{cases} \tag{21}$$

where $x^{\perp} = x - (x \cdot \hat{n})\hat{n}$ is the in-plane projection of x. These equations are the steady state solutions of the gradient flow

$$\begin{cases} \dot{\hat{n}} = -\langle (\hat{n} \cdot \nabla V) x^{\perp} \rangle_P \\ \dot{b} = \langle \hat{n} \cdot \nabla V \rangle_P, \end{cases} \tag{22}$$

where $\langle \cdot \rangle_P$ denotes the expectation with respect to $e^{-\beta V}\delta(\hat{n} \cdot x - b)$ properly normalized. Therefore (22) can be solved by evaluating the right hand-sides using blue moon sampling technique discussed in section 8.

Jóhannesson and Jónsson [JJ01] were the first to suggest optimizing the TST dividing surface among planes. However, their derivation is different from ours, and they arrive at different equations for $\hat{n}$ and b.

Whether the optimal TST dividing surface can be approximated by a plane is not *a priori* clear. However a more serious limitation of the procedure is that the optimal TST dividing surface may not solve either one of the two problems in the Bennett-Chandler procedure. Clearly the transition process may be such that the trajectories re-cross the optimal TST surface many times, i.e. the procedure does not solve the second problem in the Bennett-Chandler procedure. But it is important to realize that for such diffusive crossings the procedure may not solve the first problem either: the optimal TST surface may not be a good dividing surface, i.e. trajectories initiated from this surface may not have half-half probability to reach the metastable sets B_1 or B_2.

This indicates that the very concept of a dividing surface for the transition may be flawed. Therefore there is a need for techniques that, besides evaluating transition rates, give a more global insight about the mechanism of transition than a mere dividing surface. Techniques of this kind are discussed next.

4 Transition path sampling

In [Pra86], Pratt suggested a strategy for dealing with the rarity of transition trajectories by sampling precisely these trajectories via Monte-Carlo techniques in path space. The Transition Path Sampling (TPS) method, developed by Bolhuis, Chandler, Dellago, and Geissler, is a practical way of implementing such a strategy [DBG02, BCDG02].

We first explain TPS in the overdamped limit. The statistical weight of a trajectory $x(t)$ on $t \in [0,T]$ with $x(0) = x_0$ is, after rescaling of time by $t/\gamma \to t$, proportional to

$$\exp\Big(-\tfrac{1}{4}\beta \int_0^T |\dot{x} + \nabla V|^2 dt\Big). \tag{23}$$

In other words, any average of a functional of $x(t)$, $t \in [0,T]$, can in principle be evaluated via the path integral of this functional with the weight (23) over all paths $x(t)$ satisfying $x(0) = x_0$. There are some technical difficulties in properly normalizing (23) but since a Metropolis Monte-Carlo algorithm would only require (23) and not the normalization factor, we will leave these difficulties aside. Suppose that instead of the constraint $x(0) = x_0$ one uses $x(0) \in B_1$, $x(T) \in B_2$, then such a Monte-Carlo scheme will sample precisely the transition paths (or at least the ones that switch in less that T) with the proper weight. How to implement such a procedure in practice after time-discretization of (23) – i.e. how to generate new paths with reasonable acceptance rate in a Metropolis algorithm, etc – is explained in [DBG02, BCDG02].

In principle, TPS removes all pre-assumptions about the mechanism of transition by sampling all the paths that may be involved in such transitions. It is a brute

force calculation of the true dynamical trajectories, except that the method cleverly manages to observe these trajectories precisely during the windows of time when a transition happens. Proper analysis of the paths sampled by TPS (which is a nontrivial operation beyond TPS) may lead to better choices for potential dividing surfaces to be used in the Bennett-Chandler procedure. This removes the first difficulty mentioned in section 2. In principle, the second can also be removed by abandoning the Bennett-Chandler procedure altogether and using TPS combined with umbrella sampling in which the volume of B_2 is artificially increased until this set intersects with B_1 to compute the rate directly from TPS. This procedure is necessary because the transition rate is expressed in terms of average over path constrained on their initial position only, $x(0) \in B_1$.

TPS can be easily generalized to dynamics with arbitrary finite friction by replacing (23) by

$$\exp\Big(-\tfrac{1}{4}\beta\gamma^{-1}\int_0^T |M\ddot{x}+\gamma\dot{x}+\nabla V|^2 dt\Big). \tag{24}$$

The Hamiltonian situation is more delicate. In this case, the statistical weight on any path connecting B_1 and B_2 is either one or zero; depending on whether this path solves (5), in which case it is as good as any other such path and must be accepted, or it does not solve (5), in which case it must be rejected. Procedures have been developed to sample paths with such a weight [BDC98]. But it is not *a priori* clear how to estimate the quality of the sampling since the "energy" surface for the paths is so degenerate (constant and flat, or infinity).

As we mentioned earlier, TPS has the advantage that it makes no assumptions on the mechanism of transition. But it may be very demanding when the transition takes a relatively long time (e.g. for diffusive crossings) or when several well-separated channels for the transition exist, in which case there is also metastability in path space. In addition, analyzing the data provided by TPS, e.g. to determine a few reaction coordinates governing the transition, is nontrivial. The techniques we will discuss next provide insight about these reaction coordinates, often (but not always) at the expense of making additional approximations. These techniques are in general less expensive than TPS.

5 Identifying the minimal energy path for smooth energy landscapes

If the potential energy landscape is smooth, i.e. has no features on the thermal energy scale, and the critical points are isolated, the metastable states are the local minima and the transition states between local minima are the saddle points along the minimum energy paths (MEP) between these minima. In this section, we discuss methods for identifying MEPs.

Analytically a MEP is a heteroclinic orbit, i.e. it is a curve connecting the two minima with the property that the component of ∇V perpendicular to the curve is zero,

$$\nabla V^{\perp} = 0. \tag{25}$$

One way of finding solutions of (25) is to use the gradient flow of curves in configuration space:

$$v_n = -\nabla V^{\perp}, \tag{26}$$

where v_n is the normal velocity of the curve. There are several different ways of implementing this dynamics of curves, according to how the curves are parameterized. If $\varphi(\alpha, t)$ is a curve parameterized by $\alpha \in [0, 1]$, then the most direct implementation of (26) (proposed originally in [UE89]),

$$\frac{\partial \varphi}{\partial t} = -\nabla V^{\perp}(\varphi) = -\nabla V(\varphi) + (\nabla V(\varphi) \cdot \hat{t})\hat{t}, \qquad \hat{t} = \varphi_\alpha / |\varphi_\alpha|, \tag{27}$$

will in general not preserve the parameterization of $\varphi(\alpha, t)$ during the evolution. This may lead to computational instabilities if the discretization points along $\varphi(\alpha, t)$ happen to cluster in certain parts of the curve and leave other parts under-resolved. In [JMJ98], Jónnson, Mills, and Jacobson gave a solution to this problem by modifying (27) to

$$\frac{\partial \varphi}{\partial t} = -\nabla V(\varphi) + (\nabla V(\varphi) \cdot \hat{t})\hat{t} + \lambda(\varphi_{\alpha\alpha} \cdot \hat{t})\hat{t}, \tag{28}$$

where $\lambda > 0$ is an adjustable parameter. The new term at the right hand-side is a penalty term which helps to distribute the points uniformly along the curve after it is discretized. Note that both the potential and the spring forces are nudged in (28), respectively in the directions perpendicular and tangential to the curve. For this reason, the method is referred to as the Nudged Elastic Band (NEB) method [JMJ98].

On the other hand, the broader perspective of evolving curves with intrinsic parameterization is useful to remember since it offers possibilities different from (28) which may lead to more efficient numerical schemes. For instance instead of (28), one can use

$$\frac{\partial \varphi}{\partial t} = -\nabla V(\varphi) + (\nabla V(\varphi) \cdot \hat{t})\hat{t} + r\hat{t}, \tag{29}$$

where r is now a Lagrange multiplier for enforcing some specific parameterization of $\varphi(\alpha, t)$. A simple example is to use the equal arclength parameterization, i.e. $|\varphi_\alpha|$ is constant. But other choices are possible. A method based on (29) was introduced in [ERVE02a] under the name of the (zero-temperature) string method. The string method turns out to have several advantages. (29) is free of adjustable parameter and it naturally leads to very flexible time-splitting schemes where φ is evolved by the term $-\nabla V^{\perp}$ alone for a couple of time steps, then a re-parameterization step is performed to re-enforce proper parameterization. This makes it very simple to use any constraint on the parameterization – such as arc-length weighted by energy or curvature which put more discretization point where needed – or to change the number of discretization points at the re-parameterization step if necessary. Note that for both (28) and (29) upwind scheme for the derivative with respect to α must be used to guarantee stability of the scheme [HJ00, Ren02]. Also, higher order schemes in α can be constructed [Ren03], and convergence rate can be improved by changing the steepest descent dynamics to Broyden-like dynamics [Ren02].

Once the MEPs have been identified, the transition rate can be estimated for instance by

$$k_{12} = (4\pi)^{-1} \left(\sqrt{\gamma^2 + |\lambda_s|} - \gamma \right) \sqrt{|\det(H_m H_s^{-1})|} e^{-\beta \Delta V}, \tag{30}$$

where λ_s is the unstable eigenvalue of the Hessian of $V(x)$ evaluated at the saddle point between the two minima, H_s and H_m are the Hessians of V evaluated at the starting minimum and the saddle point, respectively, and ΔV is the energy difference between the saddle point and the starting minimum. At nonzero friction, these statements follow from the theory of large deviation and extensions thereof to account for the prefactor [FW98]. In the Hamiltonian case, (30) follows by evaluation of (11) in the harmonic approximation.

The assumptions underlying these results are very restrictive since, in many interesting situations, the energy has too many critical points. In addition these critical points are mostly irrelevant for the transitions because the saddle point(s) are not, in general, a fair approximation of the transition state. One may think that this can be checked *a posteriori* by initiating trajectories from these saddle points. But in practice this will be undoable if the number of saddle points is so large that it is impossible to identify them all.

6 Finite temperature string method

Another advantage of the (zero-temperature) string method is that it can be naturally generalized to problems with rough energy landscapes by coupling the dynamics of the string with some sampling procedure [ERVE02b]. This allows to move the string in a thermally averaged potential whose irrelevant details on the thermal scale have been smoothed out. Instead of (29) it was proposed in [ERVE02b] to use

$$\frac{\partial \varphi}{\partial t} = -\langle \nabla V(\tilde{\varphi}) \rangle + (\langle \nabla V(\tilde{\varphi}) \rangle \cdot \hat{t}) \hat{t} + r\hat{t}. \tag{31}$$

Here $\langle \nabla V(\tilde{\varphi}) \rangle$ denotes some suitably defined statistical average. A seamless way of defining such statistical averages and moving the string at the same time is given by

$$\frac{\partial \tilde{\varphi}}{\partial t} = -\nabla V(\tilde{\varphi}) + (\nabla V(\tilde{\varphi}) \cdot \hat{t}) \hat{t} + \sqrt{2\beta^{-1}} \left(\xi - (\xi \cdot \hat{t}) \hat{t} \right) + r\hat{t}. \tag{32}$$

Here $\tilde{\varphi}(\alpha, t)$ is a stochastic process whose mean is the string defined earlier in (31): $\varphi = \langle \tilde{\varphi} \rangle$. $\hat{t}$ is the unit tangent along this curve, r is a Lagrange multiplier term to enforce some constraint on the parameterization of φ, and η is a white-noise satisfying

$$\langle \xi(\alpha, t) \xi(\alpha', t') \rangle = \begin{cases} \delta(t - t') & \text{if } \alpha = \alpha' \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

Because of the resemblance of (32) with (29), the method based on (32) was called the finite temperature string method in [ERVE02b]. Notice that (31) can be obtained from averaging (32).

It is a simple matter to show that, for each α, the equilibrium density function for (32) is

$$\rho(x,\alpha) = Z^{-1}(\alpha)e^{-\beta V}\delta(\hat{t}\cdot(x-\varphi)), \tag{34}$$

i.e. it is the restriction of $e^{-\beta V}$ in the plane perpendicular to the curve φ containing the point $x = \varphi(\alpha)$, which can be parameterized as $0 = \hat{t}\cdot(x-\varphi)$. We will refer to this plane as $P(\alpha)$; note that by definition the unit normal $\hat{n}$ to this plane coincide with the unit tangent along the string, i.e.

$$\hat{n} = \hat{t}. \tag{35}$$

From (34) it follows that at statistical steady state, the string φ satisfies

$$\varphi(\alpha) = \langle x\rangle_{P(\alpha)}, \tag{36}$$

where $\langle\cdot\rangle_{P(\alpha)}$ denotes the expectation with respect to (34). Together with (35), (36) specifies completely the string φ and the associated planes P. These equations will actually be derived in section 12 in the context of adaptive sampling techniques where it will be shown that the family of planes defined by (36) emerges as a reaction coordinate to describe the transition which is optimal within a certain class. Here let us simply note that (36) is a natural finite temperature generalization of the concept of MEP. Note first that, letting $\beta\to\infty$ in (36) and assuming that the potential V is smooth, this equation reduces to

$$\varphi(\alpha) = \text{arg min } V(x) \text{ in } P(\alpha). \tag{37}$$

This equation is equivalent to (25), i.e. φ converges to a MEP as $\beta\to\infty$. At finite temperature, the path satisfying (36) can be thought of as a MEP in some thermally averaged potential in which the small features of the energy below the thermal scale have been smoothed out. This is apparent from (31). In fact, φ defines the center of a tube in configuration space whose width may for instance be characterized by the covariance matrix of $\tilde{\varphi}$:

$$\text{cov}(\tilde{\varphi}) = \left\langle (x-\varphi)(x-\varphi)^T\right\rangle_{P(\alpha)}. \tag{38}$$

The finite temperature string method averages over details of the potential within this tube and thereby accounts for entropic effects. The free energy along the path can be defined as

$$\begin{aligned} F(\alpha) &= -\beta^{-1}\log\int_{P(\alpha)} e^{-\beta V}d\sigma(x) \\ &= -\beta^{-1}\log\int_{\mathbb{R}^n} e^{-\beta V}\delta(\hat{t}\cdot(x-\varphi))dx. \end{aligned} \tag{39}$$

How to interpret this free energy and derive from it the rates for the transition will be explained in sections 11 and 12.

In practice, (32) is solved similarly as (29) except that a collection of replica of $\tilde{\varphi}$ must be evolved on each plane $P(\alpha)$, and the string φ has to be approximated by arithmetic average over these replicas. These calculations can be easily parallelized. Finally, note that in (32) the average force on the string is computed using the restricted Gibbs ensemble on the hyperplanes normal to the string. One may think of other ways of computing the averaged forces.

7 Reaction coordinates and free energy

The traditional belief in the study of rare events is that a transition between two metastable sets, B_1 and B_2, can be described by a well-chosen reaction coordinate and the free energy associated with it. In this section we give a first discussion of these concepts and we will revisit them in section 11.

A reaction coordinate is a function $q(x)$ whose level sets $q(x) = cst$ foliate the configuration space and specify the advancement of the transitions between B_1 and B_2. In the simplest setting $q(x)$ is a scalar-valued function and we will focus on this case first. The generalization to vector-valued functions is straightforward and will be discussed in the next sections. Given a reaction coordinate $q(x)$, there are two natural ways to define a free energy associated with it. The first is based on the marginal probability density in the variable q. This is the most standard free energy, and it is the one that will prove to have the right dynamical content provided that the right reaction coordinate is chosen. The second free energy is based on the probability density of the surface $q(x) = q$. This free energy is more relevant in the context of TST. Since it is important to understand the differences between them, we discuss both.

Given $q(x)$, the first natural way to define a free energy is as

$$F(q) = -\beta^{-1} \log \bar{\rho}(q), \tag{40}$$

where

$$\bar{\rho}(q) = Z^{-1} \int_{\mathbb{R}^n} e^{-\beta V(x)} \delta(q(x) - q) dx. \tag{41}$$

$\bar{\rho}$ is the marginal of the equilibrium density ρ in the variable q and (40) implies that it can also be written as $\bar{\rho}(q) = e^{-\beta F(q)}$. The introduction of F and $\bar{\rho}$ allows us to factorize the expectation of $A(x)$ with respect to $Z^{-1}e^{-\beta V}$. Indeed, $\langle A \rangle$ can be expressed as

$$\begin{aligned} \langle A \rangle &= Z^{-1} \int_{\mathbb{R}^n} A(x) e^{-\beta V(x)} dx \\ &= \int_{\mathbb{R}} \langle A \rangle_{q(x)=q} e^{-\beta F(q)} dq, \end{aligned} \tag{42}$$

where $\langle A \rangle_{q(x)=q}$ is average of $A(x)$ in the surface $q(x) = q$, with respect to $e^{-\beta V} \delta(q(x) - q)$ properly normalized. Quite remarkably, the derivative of F can expressed in terms of an average over the density $Z^{-1}e^{-\beta V(x)}$ restricted in the surface $q(x) = q$. To see this differentiate (40) with respect to q, use $|\nabla q(x)|^2 \delta'(q(x) - q) = \nabla q \cdot \nabla \delta(q(x) - q)$, and integrate by parts in x to obtain

$$F'(q) = \Big\langle \frac{\nabla q \cdot \nabla V}{|\nabla q|^2} - \beta^{-1} \nabla \cdot \Big(\frac{\nabla q}{|\nabla q|^2} \Big) \Big\rangle_{q(x)=q}. \tag{43}$$

$F'(q)$ is usually referred to as the mean force.

We will see below that the free energy in (40) is the one with the right dynamical content provided that one uses the right reaction coordinate $q(x)$. Indeed if $x(t)$ is a

trajectory in the system, the dynamics of the coarse variable $q(x(t))$ is driven by the mean force plus appropriate thermal noise. We stress however that this interpretation of $F(q)$ is different from the most direct (and also most common) one where F is mostly used to identify a good dividing surface as the surface $q(x) = q_\star$ where $F(q)$ reaches a maximum (i.e. such that the mean force vanishes on it, $F'(q_\star) = 0$). This direct interpretation is in fact incorrect, as can be seen as follows. Suppose that $q(x)$ is a reaction coordinate and let $f(z)$ be a scalar function, strictly monotonous, $f'(z) \neq 0$. Then

$$\bar{q}(x) = f(q(x)) \tag{44}$$

is as good a reaction coordinate as $q(x)$ since the level sets of both functions coincide – we shall refer to (44) as a gauge transformation. However, if one denotes by $\bar{F}(\bar{q})$ the free energy in $\bar{q}$, it is easy to see that

$$\begin{aligned}
\bar{F}(f(q)) &= -\beta^{-1} \log \int_{\mathbb{R}^n} e^{-\beta V(x)} \delta(\bar{q}(x) - f(q)) dx + \beta^{-1} \log Z \\
&= -\beta^{-1} \log \int_{\mathbb{R}^n} e^{-\beta V(x)} (f'(q(x)))^{-1} \delta(q(x) - q) dx + \beta^{-1} \log Z \\
&= F(q) + \beta^{-1} \log f'(q),
\end{aligned} \tag{45}$$

i.e. $\bar{F}(f(q)) \neq F(q)$, the free energy is not left invariant by the gauge transformation in (44). In particular, if $q_\star$ is a critical point (minimum or maximum) of $F(q)$, i.e. $F'(q_\star) = 0$, one has

$$\bar{F}'(f(q_\star)) = \beta^{-1} f''(q_\star)/(f'(q_\star))^2 \neq 0, \tag{46}$$

which is not equal to zero in general. This is a serious flaw of the argument which identifies the critical points of the free energy in (40) as metastable sets or transition states.

The second natural way of defining a free energy is as follows (compare (40))

$$\begin{aligned}
G(q) &= -\beta^{-1} \log Z^{-1} \int_{q(x)=q} e^{-\beta V(x)} d\sigma(x) \\
&= -\beta^{-1} \log Z^{-1} \int_{\mathbb{R}^n} e^{-\beta V(x)} |\nabla q(x)| \delta(q(x) - q) dx.
\end{aligned} \tag{47}$$

With this definition $e^{-\beta G(q)}$ is the probability density of the surface $q(x) = q$, and $G(q)$ is indeed gauge invariant: $\bar{G}(f(q)) = G(q)$ if $\bar{q} = f(q)$ and $\bar{G}$ is the free energy in $\bar{q}$. $G(q)$ is related to the standard free energy $F(q)$ defined in (40) via the relation

$$e^{-\beta G(q)} = \langle |\nabla q| \rangle_{q(x)=q} e^{-\beta F(q)} \tag{48}$$

Therefore it is easy to see that with this choice one has (compare (42))

$$\langle A \rangle = \int \frac{\langle A \rangle_{q(x)=q}}{\langle |\nabla q| \rangle_{q(x)=q}} e^{-\beta G(q)} dq. \tag{49}$$

Also the mean force associated with G can be expressed as the following conditional average:

$$G'(q) = \left\langle \hat{n} \cdot \nabla V - \beta^{-1} \kappa \right\rangle_{q(x)=q}, \tag{50}$$

where $\hat{n} = \nabla q / |\nabla q|$ is the unit normal to the surface $q(x) = q$, and $\kappa = \nabla \cdot \hat{n}$ is the local mean curvature of this surface. Quite interestingly, this formula implies that $G'(q_\star) = 0$ if

$$0 = \left\langle \hat{n} \cdot \nabla V - \beta^{-1} \kappa \right\rangle_{q(x)=q_\star} \tag{51}$$

This equality is an averaged version of equation (18) for the optimal TST surface. In fact it is easy to show that if one optimizes the object function in (17) over the level sets of a given reaction coordinate $q(x)$, then the optimal TST dividing surface is the level set of $q(x)$ where (51) is satisfied, i.e. it is a critical point of the free energy $G(q)$ (actually it can be shown that it is a maximum by checking that $G''(q_\star) < 0$). This indicates that the definition in (47) for the free energy is the natural one in the context of TST.

We conclude this section by a severe warning. As mentioned before, one usually hopes that the reaction coordinate describes the advancement of the transitions between B_1 and B_2, i.e. that it has some dynamical meaning which can be deduced by analyzing the properties of the free energy $F(q)$ or $G(q)$. But clearly this requires to pick the right reaction coordinate since the characteristics of $F(q)$ and $G(q)$ depend sensitively on $q(x)$. The results in this section leave completely open the question of how to choose $q(x)$. This question is highly nontrivial, particularly since the order parameter used to distinguish the metastable basins may be a bad reaction coordinate for the transition between these states. It is also particularly important since knowing the right reaction coordinate often means that we have a good intuitive understanding of how the reaction proceeds. How to choose $q(x)$ will be discussed in section 11.

8 Blue-moon sampling technique

It is remarkable that one can actually compute rather efficiently the free energy in (40) and the conditional expectation in (42) by using constrained simulations. This was first noted by Carter, Ciccotti, Hynes, and Kapral in [CCHK89] (see also [SC98]) where the blue-moon sampling technique was introduced. The idea is to use ergodicity and evaluate the conditional average in (43) for $F'(q)$ by time-averaging over a trajectory whose dynamics is constrained such that $q(x) = q$ and has as its equilibrium density $Z^{-1} e^{-\beta V(x)}$ restricted in this surface. F can then be estimated by integration of F' – a step referred to as thermodynamic integration [FS01]. In [CCHK89], it was proposed to compute the expectation in (43) via Hamiltonian dynamics simulations subject to proper constraint. Another possibility is to use

$$\dot{z} = -(\nabla V(z))^\perp - \beta^{-1} \nabla \cdot (\hat{n} \otimes \hat{n}) + \sqrt{2\beta^{-1}}\, \eta^\perp \tag{52}$$

where η is a white-noise, $a^\perp = a - (a \cdot \hat{n})\hat{n}$ with $\hat{n} = \nabla q / |\nabla q|$ denotes the projection of a into the hypersurface defined by $q(x) = q$. The term $-\beta^{-1} \nabla \cdot (\hat{n} \otimes \hat{n})$ is a

spurious drift term that arises since the products in $\eta^\perp$ are interpreted in Itô sense. It is easy to check that (52) has the right equilibrium density, i.e. if $q(z(0)) = q$, then

$$\langle A \rangle_{q(x)=q} = \lim_{T\to\infty} \frac{1}{T} \int_0^T A(z(t))dt, \tag{53}$$

offers a practical way to evaluate the conditional expectations in (42) and (43). (52) can be solved using a time-splitting scheme such as

$$\begin{aligned} z^\star &= z^n - (\nabla V(z^n))^\perp \Delta t + \sqrt{2\beta^{-1}}\,(\eta^n)^\perp \\ z^{n+1} &= z^\star + \lambda \nabla \cdot (\hat{n}(z^n) \otimes \hat{n}(z^n)). \end{aligned} \tag{54}$$

Here Δt is the time-step, the η^n's are independent identically distributed Gaussian random variables with mean zero and variance Δt, and λ is a scalar quantity such that

$$q(z^{n+1}) = q(z^n). \tag{55}$$

A scheme based on (54) and (55) is very much in the spirit of the method SHAKE introduced in [RCB77] (see also [WCH01]).

It is also interesting to note that the blue moon procedure can be generalized to multi-dimensional reaction coordinates where one uses a set of function $q_1(x)$, $\ldots$, $q_m(x)$ instead of a single $q(x)$. The associated free energy $F(q_1, \ldots q_m)$ is the multidimensional analog of (40)

$$F(q_1, \ldots q_m) = -\beta^{-1} \log Z^{-1} \int_{\mathbb{R}^n} e^{\beta V(x)} \delta(q_1(x) - q_1) \cdots \delta(q_m(x) - q_m) dx \tag{56}$$

If the functions $q_1(x)$,$\ldots$, $q_m(x)$ satisfy

$$\nabla q_j \cdot \nabla q_k = 0 \qquad \text{if } j \neq k, \tag{57}$$

i.e. they form an orthogonal set of curvilinear coordinates, then the gradient of the free energy in $q_1, \ldots, q_m$ can be expressed as

$$\frac{\partial F}{\partial q_j} = \Big\langle \frac{\nabla q_j \cdot \nabla V}{|\nabla q_j|^2} - \beta^{-1} \nabla \cdot \Big(\frac{\nabla q_j}{|\nabla q_j|^2} \Big) \Big\rangle_{q_1(x)=q_1, \ldots, q_m(x)=q_m}. \tag{58}$$

This average can be obtained via constrained simulations, e.g. similar to (52). Of course, retrieving $F(q_1, \ldots, q_m)$ from its gradient (58) will be more complicated than in the one-dimensional case. One possibility is to use methods such as NEB or the zero-temperature string method that only requires (58) as its input to at least determine saddle points and heteroclinic orbits on the free energy surface (that is, the minimum free energy paths, MFEPs). Since the free energy surface will in general be much smoother than the original potential, it will have much less critical points, and MFEPs may be relevant even in situations where MEPs are not.

9 Metadynamics

At the end of section 8 we suggested how to determine the MFEPs in a two-step procedure that would combine the blue moon sampling technique with methods like NEB or the zero temperature string method. Recently in [ILP03] (see also [LP02]), Iannuzzi, Laio, and Parrinello proposed an alternative technique, termed metadynamics, which permits to sample in a seamless way the free energy surface in a set of prescribed reaction coordinates $q_1(x), \ldots, q_m(x)$. The idea is to extend the phase-space so as to include the reaction coordinates as additional dynamical variables. [ILP03] works in the Hamiltonian context, but we will consider the case of overdamped dynamics which is better suited for analysis.

Metadynamics amounts to considering the thermally perturbed gradient flow on the extended energy

$$U(x, Q) = V(x) + \tfrac{1}{2}\mu|Q - q(x)|^2, \tag{59}$$

where $\mu > 0$ is a parameter to be prescribed later, and $Q = (Q_1, \ldots, Q_m)$ are the additional variables corresponding to the reaction coordinates. Associated with (59) we take the following equations of motion:

$$\begin{cases} \dot{x} = -\delta\nabla V(x) + \delta\mu(Q - q(x))\nabla q(x) + \sqrt{2\beta^{-1}\delta}\,\eta, \\ \dot{Q} = -\mu(Q - q(x)) + \sqrt{2\beta^{-1}}\,\eta_q, \end{cases} \tag{60}$$

where η_q is a white-noise independent of η, and $\delta > 0$ is another adjustable parameter. We will consider the case when $\delta > 1$ corresponding to situations where there is more friction on $Q(t)$ than on $x(t)$ – i.e. $Q(t)$ are slower variables than $x(t)$.

For all $\delta > 0$, the equilibrium density function for (60) is

$$\rho_\mu(x, Q) = Z^{-1}(\mu\beta/2\pi)^{m/2} e^{-\beta V(x) - \frac{1}{2}\mu\beta|Q - q(x)|^2}, \tag{61}$$

where Z is the normalization factor of $e^{-\beta V(x)}$ alone, $Z = \int_{\mathbb{R}^n} e^{-\beta V(x)} dx$. The corresponding marginal density on Q is

$$\bar{\rho}_\mu(Q) = Z^{-1}(\mu\beta/2\pi)^{m/2} \int_{\mathbb{R}^n} e^{-\beta V(x) - \frac{1}{2}\mu\beta|Q - q(x)|^2} dx \tag{62}$$

Viewed as a function of Q, $(\mu\beta/2\pi)^{m/2} e^{-\frac{1}{2}\mu\beta|Q - q(x)|^2}$ converges weakly towards $\delta(Q - q(x))$ as $\mu \to \infty$. Therefore

$$\bar{\rho}_\mu(q) \to \bar{\rho}(q) \qquad \text{as } \mu \to \infty, \tag{63}$$

where $\bar{\rho}(q)$ is the multi-dimensional analog of the marginal density in the variables q that we defined in (41). A similar argument shows that (62) is also the reduced density for the extended Hamiltonian system

$$\begin{cases} M\ddot{x} = -\nabla V(x) + \mu(Q - q(x))\nabla q(x) \\ M_q\ddot{Q} = -\mu(Q - q(x)), \end{cases} \tag{64}$$

corresponding to the extended Lagrangian

$$L(x,Q) = \tfrac{1}{2}(\dot{x}, M\dot{x}) - V(x) + \tfrac{1}{2}(\dot{Q}, M_q\dot{Q}) - \tfrac{1}{2}\mu|Q-q(x)|^2. \tag{65}$$

Here M_q is the mass tensor associated with the new variable Q. [ILP03] works in the Hamiltonian context and use (64), but we will stick to (60).

The idea in [ILP03] is to use (64) at large μ and large M_Q (with μ/M_Q small) to compute the free energy associated with $\bar{\rho}(q)$. Here we show that this free energy can also be computed from (60) at large μ and large δ with μ/δ small. When δ is large, the variables $Q(t)$ evolves much more slowly than $x(t)$ and only feel the average effect of the latter. The proper conditional probability density with which to average the right hand-side of (60) is (compare (61))

$$\rho_\mu(x|Q) = Z_\mu^{-1}(Q) e^{-\beta V(x) - \frac{1}{2}\mu\beta|Q-q(x)|^2}, \tag{66}$$

where

$$Z_\mu(Q) = \int_{\mathbb{R}^n} e^{-\beta V(x) - \frac{1}{2}\mu\beta|Q-q(x)|^2} dx, \tag{67}$$

is a normalization factor which guarantees that $\int_{\mathbb{R}^n} \rho_\mu(x|Q)dx = 1$ for all Q. Since

$$\begin{aligned} &- Z_\mu^{-1}(Q) \int_{\mathbb{R}^n} \mu(Q-q(x)) e^{-\beta V(x) - \frac{1}{2}\mu\beta|Q-q(x)|^2} dx \\ &\quad = \beta^{-1} Z_\mu^{-1}(Q) \nabla_Q Z_\mu(Q) \\ &\quad = \beta^{-1} \nabla_Q \log Z_\mu(Q), \end{aligned} \tag{68}$$

the limiting equations for $Q(t)$ (60) as $\delta \to \infty$ can be written as

$$\dot{Q} = -\nabla_Q F_\mu + \sqrt{2\beta^{-1}}\,\eta_q, \tag{69}$$

provided one defines

$$F_\mu(Q) = -\beta^{-1} \nabla_Q \log Z_\mu(Q). \tag{70}$$

This is a mollified version of the free energy since (67) and (70) imply that

$$F_\mu(Q) \to F(Q) \tag{71}$$

as $\mu \to \infty$ and therefore the limiting equation for $Q(t)$ as $\delta \to \infty$ and $\mu \to \infty$ (in this order) is

$$\dot{Q} = -\nabla_Q F + \sqrt{2\beta^{-1}}\,\eta_q, \tag{72}$$

Thus, by simulating (60) at large δ, μ, and monitoring the evolution of $Q(t)$ one can indeed sample the free energy landscape in the variables $q(x)$. But there is still one difficulty: (60) will display metastability if the original dynamics does, so this equation may not be practical. To fix this problem, Iannuzzi, Laio, and Parrinello suggest to use a technique introduced in [HTV94] and [WL01] and further modify the dynamics by including in (60) (or rather (64) in the original paper) an additional

non-Markovian term which discourages the trajectory from going back to regions that it has already visited. For instance, one may modify (60) as

$$\begin{cases} \dot{x} = -\delta \nabla V(x) + \delta \mu (Q - q(x)) \nabla q(x) + \sqrt{2\beta^{-1}\delta}\, \eta, \\ \dot{Q} = -\mu (Q - q(x)) + \sqrt{2\beta^{-1}}\, \eta_q \\ \qquad\quad + A \displaystyle\int_0^t (Q(t) - Q(t')) e^{-|Q(t)-Q(t')|^2/\Delta q^2} dt' \end{cases} \tag{73}$$

where A and Δq are adjustable parameters. Proceeding as before, it is easy to see that the limiting equation for Q as $\delta \to \infty$ and $\mu \to \infty$ is

$$\dot{Q} = -\nabla_Q F + A \int_0^t (Q(t) - Q(t')) e^{-|Q(t)-Q(t')|^2/\Delta q^2} dt' + \sqrt{2\beta^{-1}}\, \eta_q, \tag{74}$$

Therefore, the memory term added in (73) is a term that fills up the potential well that the trajectory has already visited. In particular, if

$$U(q,t) = \tfrac{1}{2} A \Delta q^2 \int_0^t e^{-|q-Q(t')|^2 \Delta q^2} dt', \tag{75}$$

then, as $t \to \infty$, $U(q,t) - U(q',t)$ converges to an estimate of $F(q') - F(q)$ when $\delta \gg \mu \gg 1$. The parameters A and Δq control the accuracy on the resolution of the free energy: as they are decreased, the resolution improves, but the convergence rate in time deteriorates. Given some accuracy requirement, estimating the optimal choice of parameters δ, μ, A, and Δq in a metadynamics calculation is a nontrivial question which we will leave aside.

10 Another type of metadynamics: the accelerated dynamics

The metadynamics calculation of section 9 gives the free energy of the set of reaction coordinates $q(x) = (q_1(x), \ldots, q_m(x))$ in the specific gauge fixed by this choice. We wish now to propose an alternative type of metadynamics – the accelerated dynamics – which shed some light on the issue of choice of gauge. Given a set of reaction coordinates $q(x)$, consider

$$\begin{aligned} \dot{x} = {} & \delta P(-\nabla V(x) + \sqrt{2\beta^{-1}}\, \eta) \\ & + (1-P)(-\nabla V(x) + \sqrt{2\beta^{-1}}\, \eta) + (\delta - 1)\beta^{-1} \nabla \cdot P, \end{aligned} \tag{76}$$

where

$$P = 1 - \sum_{j=1}^{m} \hat{n}_j \otimes \hat{n}_j, \qquad \hat{n}_j = \nabla q_j / |\nabla q_j|. \tag{77}$$

P is the projector into the surface where $q(x) = cst$ (here for simplicity we assume that (57) holds). When $\delta = 1$, (76) reduces to the overdamped equation (4) (in time

units where $\gamma = 1$). When $\delta > 1$, (76) corresponds to a dynamics where one has artificially accelerated the dynamics on $q(x) = cst$ while keeping its original speed in the direction perpendicular to these surfaces. The term $(\delta - 1)\beta^{-1}\nabla \cdot P$ is again a spurious drift term from the interpretation of the noise terms in (76) in Itô sense. A set of equations analogous to (76) can be written in the Hamiltonian setting,

$$\begin{cases} \dot{x} = \delta P v + (1 - P)v \\ M\dot{v} = -\delta P \nabla V(x) - (1 - P)\nabla V(x), \end{cases} \tag{78}$$

but we will again stick to (76) as it is better suited for analysis. The equilibrium density function of both (76) and (78) is

$$\rho(x) = Z^{-1} e^{-\beta V(x)} \tag{79}$$

for all values of $\delta > 0$.

Let $Q(t) = q(x(t))$. It is easy to see that $Q(t)$ satisfies

$$\dot{Q} = -\nabla V \cdot \nabla q + \beta^{-1} \Delta q + \sqrt{2\beta^{-1}}(\eta \cdot \nabla) q. \tag{80}$$

(80) is not closed. But if we now assume that $\delta \gg 1$, the motion on the surface $q(x) = cst$ can be averaged out. It is shown in Appendix B that in the limit as $\delta \to \infty$ (80) reduces to the following closed equation for $Q(t)$:

$$\dot{Q}_j = -a_j(Q)\frac{\partial F}{\partial Q_j} + \beta^{-1}\frac{\partial a_j}{\partial Q_j} + \sqrt{2\beta^{-1}a_j(Q)}\,\eta_j, \quad j = 1, \ldots, m. \tag{81}$$

Here

$$F(q) = -\beta^{-1} \log Z^{-1} \int_{R^n} e^{-\beta V} \delta(q_1 - q_1(x)) \ldots \delta(q_m - q_m(x)) dx, \tag{82}$$

is the free energy in the variables q and

$$a_j(q) = \langle |\nabla q|^2 \rangle_{q(x)=q} \tag{83}$$

where $\langle \cdot \rangle_{q(x)=q}$ denotes expectation in $q(x) = q$ with respect to $e^{-\beta V}$ properly normalized. Since the forward operator associated with (81) is

$$L_q \rho(q) = \sum_{j=1}^{m} \frac{\partial}{\partial q_j}\Big(a_j(q)\frac{\partial F}{\partial q_j}\rho + \beta^{-1} a_j(q)\frac{\partial \rho}{\partial q_j}\Big) \tag{84}$$

it is easy to see that $e^{-\beta F}$ is the equilibrium density for (81).

In general, (81) is not in the standard form (72) since $a_j \neq 1$. But in contrast to what happens with (72), the dynamics associated with (81) is left invariant by a gauge transformation. Consider the multidimensional analog of (44)

$$\bar{q}_j(x) = f_j(q_j(x)), \quad j = 1, \ldots, m. \tag{85}$$

$\bar{q}(x)$ and $q(x)$ have the same level sets, i.e. they define the same reaction coordinate. And it is easy to see that the limiting equation similar to (81) one obtains for $\bar{Q}(t) = \bar{q}(x(t))$ guarantees that

$$\bar{Q}_j(t) = f_j(Q_j(t)). \tag{86}$$

This is in fact obvious since (76) is gauge invariant. In addition if

$$a_j(q) = 1, \tag{87}$$

then (81) takes the standard form

$$\dot{Q}_j = -\frac{\partial F}{\partial Q_j} + \sqrt{2\beta^{-1}}\,\eta_j, \quad j = 1, \ldots, m. \tag{88}$$

(87) can always be satisfied with suitable gauge transformation (85) on $q(x)$ such that

$$\sqrt{a_j(q)}\, f_j'(q) = 1, \qquad j = 1, \ldots, m. \tag{89}$$

These considerations suggest that given the appropriate set of reaction coordinates $q(x) = (q_1(x), \ldots, q_m(x))$ (but regardless of the choice of gauge), the limiting accelerated dynamics in (81) may still capture some feature of the original, unaccelerated dynamics in (4). In section 11 we will show that this is indeed the case and that the transition rates between metastable basins are independent of the value of $\delta \geq 1$ – and in particular are captured by the limiting equation in (81) – provided that the right reaction coordinates are used.

Finally, we note that (76) can in principle be used to compute the free energy by using a flooding procedure similar to one used in [LP02, ILP03]. The idea is to modify (76) as

$$\begin{aligned}\dot{x} = {} & \delta P(-\nabla V(x) + \sqrt{2\beta^{-1}}\,\eta) \\ & + (1-P)(-\nabla V(x) + \sqrt{2\beta^{-1}}\,\eta) + (\delta - 1)\beta^{-1}\nabla \cdot P \\ & + A\sum_{j=1}^{m} \nabla q_j(x) \int_0^t (q_j(x(t)) - q_j(x(t')))e^{-|q(x(t)) - q(x(t'))|^2/\Delta q^2} dt',\end{aligned} \tag{90}$$

where A and Δq are adjustable parameters similar to the ones in (74). As $\delta \to \infty$, (90) leads to the following limiting equation for $Q(t)$

$$\begin{aligned}\dot{Q}_j = {} & -a_j(Q)\frac{\partial F}{\partial Q_j} + \beta^{-1}\frac{\partial a_j}{\partial Q_j} + \sqrt{2\beta^{-1}a_j(Q)}\,\eta_j \\ & + Aa_j(Q)\int_0^t (Q_j(t) - Q_j(t'))e^{-|Q(t) - Q(t')|^2/\Delta q^2} dt',\end{aligned} \tag{91}$$

or, in the gauge where $a_j(q) = 1$,

$$\dot{Q}_j = -\frac{\partial F}{\partial Q_j} + \sqrt{2\beta^{-1}}\,\eta_j + A\int_0^t (Q_j(t) - Q_j(t'))e^{-|Q(t) - Q(t')|^2 \Delta q^2} dt'. \tag{92}$$

It follows that we can estimate the free energy by monitoring the evolution of

$$U(q,t) = \tfrac{1}{2} A \Delta q^2 \int_0^t e^{-|q-q(x(t'))|^2/\Delta q^2} dt', \tag{93}$$

since $U(q,t) - U(q',t)$ converges as $t \to \infty$, to an estimate of $F(q') - F(q)$ when $\delta \gg 1$.

11 Choosing the right reaction coordinates

The main limitation of the blue moon sampling technique discussed in section 8 or the method based on metadynamics discussed in sections 9 and 10 is that the reaction coordinate $q(x)$ must be specified beforehand. Here we discuss how to get around this problem and how to determine "optimal" reaction coordinates. We consider first the overdamped dynamics in (4) and then we will indicate how to generalize the concepts to other types of dynamics.

We start by giving a more precise definition of metastability. A system will be metastable with respect to the dynamics in (4) if the eigenvalues of the backward operator associated with this equation contain a spectral gap, i.e. $\lambda_1/\lambda_2 \ll 1$ where λ_1 is the smallest nonzero eigenvalue of $L = -\nabla V \cdot \nabla + \beta^{-1}\Delta$ and λ_2 is the second smallest eigenvalue. Clearly, this definition encompasses the concept of separation of time scales between the relaxation time within the sets B_1 and B_2, and the transition time between these sets. In fact it is well known that λ_1 is a good approximation of the relaxation rate,

$$\lambda_1 \approx k_{12} + k_{21}. \tag{94}$$

The error in this expression is $O(\lambda_1/\lambda_2)$. (94) combined with the relation $N_1/N_2 = k_{21}/k_{12}$ allows us to determine the rates.

It is also known [HMS02, SH02, BEGK03a, BEGK03b] that, if $\lambda_1/\lambda_2 \ll 1$, we have

$$\lambda_1 \approx Z^{-1} \frac{N_1 + N_2}{N_1 N_2} \int_\Omega |\nabla q|^2 e^{-\beta V} dx \tag{95}$$

where $\Omega = \mathbb{R}^n/(B_1 \cup B_2)$ and $q(x)$ is the solution of the backward equation

$$0 = -\nabla V \cdot \nabla q + \beta^{-1} \Delta q \tag{96}$$

in Ω with Dirichlet boundary conditions $q|_{\partial B_1} = 0$, $q|_{\partial B_2} = 1$. (95) holds because the eigenfunction $\varphi_1(x)$ associated with $\lambda_1 = \int_{\mathbb{R}^n} |\nabla \varphi_1|^2 e^{-\beta V} dx$ is approximately constant over regions that contain B_1 and B_2 and can be approximated by q appropriately rescaled as in (95).

In addition to (95) the solution q of (96) has the following remarkable properties which qualify it as an "optimal" reaction coordinate. First the level sets of $q(x)$ are the isoprobability surfaces for the transition between B_1 and B_2 because

$$q(x) = \text{Prob}\{X(t,x) \text{ reaches } B_2 \text{ before } B_1\}, \tag{97}$$

where $X(t,x)$ denotes the solution of (4) with initial condition $X(0,x)=x$. Since a reaction coordinate is supposed to describe the advancement of the transition between B_1 and B_2 at the coarse-grained level where all the information within the level set $q(x)=cst$ is averaged out, the function $q(x)$ satisfying (97) is arguably as good as it can get with any reaction coordinate.

Because they are isoprobability surfaces, the level sets of $q(x)$ have also the property that the probability density of the transition paths between B_1 and B_2 restricted to the surface $q(x)=q$ is precisely the equilibrium density $e^{-\beta V(x)}$ restricted to this surface. Assume that the equilibrium density on each of these surfaces is localized, and let T be a tube in configuration space such that the intersection of this tube with each surface $q(x)=q$ supports most of the probability on this surface, i.e. such that

$$\int_T e^{-\beta V(x)}\delta(q(x)-q)dx \approx \int_{\mathbb{R}^n} e^{-\beta V(x)}\delta(q(x)-q)dx \tag{98}$$

for each q. Then with probability close to one the transition paths between B_1 and B_2 stay in T. In fact, both the level set of $q(x)$ and the tube T should be used to characterize the reaction coordinate of the transitions between B_1 and B_2. Note that multiple tubes instead of a single one may have to be introduced when there are more than one channel for the transition. The generalization to this case is straightforward.

Another nice property of the solution of (96) offers possible ways of generalizing the concept to other type of dynamics, or to use more than one reaction coordinate. As shown in Appendix C, the relaxation rate in (95) is precisely the one which one obtains from the limiting accelerated equation in (81). But since the solution of (96) is also the minimizer of

$$I_1 = \int_\Omega |\nabla q|^2 e^{-\beta V} dx, \tag{99}$$

this indicates that, among all $q(x)$, the solution of (96) is the one that minimizes the transition rate of the accelerated dynamics (in which case it is in fact the transition rate of the original, un-accelerated dynamics to order $O(\lambda_1/\lambda_2)$). This property should be generic and independent of the type of dynamics or the number of reaction coordinates one chooses to describe the transition (though, in the present case, (95) clearly indicates that a scalar-valued $q(x)$ is enough). We believe it could be taken as a starting point for determining adaptively optimal reaction coordinates for arbitrary dynamics by generalizing the numerical procedure for identifying $q(x)$ and T which we discuss next.

12 Adaptive sampling techniques

We now turn the observations of section 11 into practical tools to determine the "optimal" reaction coordinate and the transition tube. These can be viewed as a way to do adaptive blue moon sampling.

Besides being the minimizer of I_1, q is also the maximizer of

$$\begin{aligned} I_2 &= \int_0^1 \Big(\int_{q(x)=q} |\nabla q| e^{-\beta V} d\sigma(x) \Big)^{-1} dq \\ &= \int_0^1 \Big(\int_{\mathbb{R}^n} |\nabla q|^2 e^{-\beta V} \delta(q(x)-q) dx \Big)^{-1} dq, \end{aligned} \tag{100}$$

with $\sup_q I_2 = (\inf_q I_1)^{-1}$. This is shown in Appendix C. (100) has the advantage of being invariant under the gauge transformation in (44), $q(x) \to \bar{q}(x) = f(q(x))$. Note also that in the gauge when $\langle |\nabla q|^2 \rangle_{q(x)=q} = 1$ (i.e. when $a = 1$ and the alternative metadynamics limiting equation is in the standard form (88)), (100) can be written as

$$I_2 = \int_0^1 e^{\beta F(q)} dq. \tag{101}$$

To turn these observations into a computational procedure for determining the optimal reaction coordinate, we can maximize the functional I_2 over specific classes of $q(x)$. The simplest choice is to assume that the level sets of $q(x)$ are locally planar in the tube T specified by (98). It is convenient to specify these planes by a parameterized curve – i.e. a string – $\varphi(\alpha)$, with $\alpha \in [0,1]$, such that (i) the plane labeled by α, which we will denote by $\bar{P}(\alpha)$ contains the point $x = \varphi(\alpha)$, (ii) the unit normal of $\bar{P}(\alpha)$ is $\hat{n}(\alpha)$, and (iii) the point $\varphi(\alpha)$ is the mean position in the plane, i.e.

$$\varphi(\alpha) = \langle x \rangle_{\bar{P}(\alpha)}, \tag{102}$$

where $\langle \cdot \rangle_{\bar{P}(\alpha)}$ denotes the expectation with respect to $e^{-\beta V(x)} \delta(\hat{n} \cdot (x - \varphi))$ properly normalized. The localization approximation requires that

$$(\hat{n} \cdot \varphi_\alpha)^2 \gg \big\langle (\hat{n}_\alpha \cdot (x - \varphi))^2 \big\rangle_{\bar{P}(\alpha)} \tag{103}$$

(103) relates the width of the transition tube T to the local curvature of the string and it is required since otherwise the probability on the planes would include regions where these planes intersect, thereby invalidating the local planar assumption for the level sets of $q(x)$. When (103) is satisfied, it is shown in Appendix C that (100) reduces to

$$\begin{aligned} I_2 &= \int_0^1 \Big(\int_{\bar{P}(\alpha)} e^{-\beta V} d\sigma(x) \Big)^{-1} \hat{n} \cdot \varphi_\alpha d\alpha \\ &= \int_0^1 \Big(\int_{\mathbb{R}^n} e^{-\beta V} \delta(\hat{n} \cdot (x - \varphi)) dx \Big)^{-1} \hat{n} \cdot \varphi_\alpha d\alpha, \end{aligned} \tag{104}$$

subject to (102). Furthermore, the curve maximizing this functional satisfies

$$0 = \langle (\hat{n} \cdot \nabla V)(x - \varphi) \rangle_{\bar{P}(\alpha)} \tag{105}$$

Comparing (105) with (22) we see that the plane along the string where $\langle \hat{n} \cdot V \rangle_{\bar{P}} = 0$ coincides with the optimal TST plane. Furthermore, within the approximation in (103), it is shown in Appendix C that (105) is equivalent to

$$\varphi_\alpha / |\varphi_\alpha| = \hat{n}, \tag{106}$$

(106) indicates that the unit tangent $\hat{t} = \varphi_\alpha / |\varphi_\alpha|$ along the string coincides with the unit normal $\hat{n}$ to the planes. It is remarkable that (106) and (102) are the same set of conditions as (36) and (35). In other words, under the localization assumption, the planes perpendicular to the string determined in a seamless way by the finite temperature string method coincide with the level sets of the optimal reaction coordinate determined by optimization of (100) within the class of $q(x)$ with (locally) planar level sets.

13 Concluding remarks

We have reviewed a variety of techniques for determining transition pathways and rates in complex systems. We left aside many important techniques such as hyperdynamics [SV00, Vot97], stochastic difference equation [EGC02, ECGS03], or clustering algorithms based on spectral analysis of the Perron-Frobenius operator [FSDC02, SHD02]. Most of the techniques we discussed require prior knowledge of the final states for the transition, but we have also included techniques such as metadynamics which does not require such information. Some of these techniques such as NEB and the zero-temperature string method are predicated on the notion of MEP, and therefore are only useful when the transition proceeds through isolated saddle points of the potential energy landscape. Others, such as TPS, blue moon sampling and the finite-temperature string method, should in principle work even when the energy landscape is rather rough with a dense set of critical points, for which the notion of MEP is no longer relevant. Some of the techniques such as blue moon sampling and metadynamics require specifying the reaction coordinates beforehand. Others, such as TPS, the finite-temperature string method and adaptive sampling techniques, do not require knowing the reaction coordinates beforehand.

From a conceptual point of view, the classical TST is based on one dividing surface. Consequently it is not able to give an accurate description of diffusive barrier crossing. The Bennett-Chandler procedure introduces two more surfaces, ∂B_1 and ∂B_2 which are the boundaries of the metastable sets. This gives a well-defined notion of transmission coefficient but the transmission coefficient may be hard to evaluate in practice because of the difficulties mentioned in section 2. These difficulties can in principle be overcome by introducing the optimal foliation between the two metastable sets. We believe that practical techniques can be developed using this framework, along the lines discussed in section 11.

We have seen that a fairly coherent theoretical framework can be developed for the high friction limit. The main remaining theoretical difficulty is associated with pure Hamiltonian dynamics. This is the case of most practical interest. We hope that some of the notions reviewed here can at least serve as a starting point for developing approximations for transition pathways and transition rates in that case.

Acknowledgments

In our understanding of the various techniques we reviewed we benefited from discussions with the authors of these techniques. We thank them all. We are especially grateful to Giovanni Ciccotti and Ray Kapral for their comments about the material in sections 3 and 11. The work of E is supported in part by ONR grant N00014-01-1-0674. The work of Vanden-Eijnden is supported in part by NSF grants DMS01-01439, DMS02-09959 and DMS02-39625.

A Derivation of (18) and (21)

Suppose that the dividing surface ∂A is parameterized as $0 = q(x)$, i.e. it is the zero level set of some function $q(x)$. Then (17) can be written as

$$I_0 = \int_{\mathbb{R}^n} |\nabla q| e^{-\beta V} \delta(q(x)) dx. \tag{107}$$

Denote by δq the variation in q, the first variation of the above functional with respect to $q(x)$ is

$$\delta I_0 = \int_{\mathbb{R}^n} e^{-\beta V(x)} \hat{n} \cdot \nabla \delta q \, \delta(q(x)) dx + \int_{\mathbb{R}^n} e^{-\beta V(x)} |\nabla q| \delta q \, \delta'(q(x)) dx. \tag{108}$$

where $\hat{n} = \nabla q / |\nabla q|$ is the unit normal of $q(x) = 0$. Since

$$\delta'(q(x)) = |\nabla q|^{-1} \hat{n} \cdot \nabla \delta(q(x)), \tag{109}$$

we obtain after integration by parts for the second integral in (108)

$$\begin{aligned}
\delta I_0 &= \int_{\mathbb{R}^n} e^{-\beta V(x)} \hat{n} \cdot \nabla \delta q \, \delta(q(x)) dx + \int_{\mathbb{R}^n} e^{-\beta V(x)} \delta q \, \hat{n} \cdot \nabla \delta(q(x)) dx \\
&= \int_{\mathbb{R}^n} e^{-\beta V(x)} \hat{n} \cdot \nabla \delta q \, \delta(q(x)) dx \\
&\quad + \int_{\mathbb{R}^n} e^{-\beta V(x)} \Big(\big(\beta \nabla V \cdot \hat{n} - \nabla \cdot \hat{n} \big) \delta q - \hat{n} \cdot \nabla \delta q \Big) \delta(q(x)) dx \\
&= \int_{\mathbb{R}^n} e^{-\beta V(x)} \big(\beta \nabla V \cdot \hat{n} - \nabla \cdot \hat{n} \big) \delta q \, \delta(q(x)) dx.
\end{aligned} \tag{110}$$

Therefore, on $q(x) = 0$, the minimizer of I_0 satisfies (18).

The derivation of (21) from (20) is similar. The first variation of this functional with respect to $\hat{n}$ and b gives

$$\begin{aligned}
\delta I &= \int_{\mathbb{R}^n} e^{-\beta V} (\delta \hat{n}^{\perp} \cdot x - \delta b) \delta'(\hat{n} \cdot x - b) dx \\
&= \int_{\mathbb{R}^n} e^{-\beta V} (\delta \hat{n} \cdot x^{\perp} - \delta b) \hat{n} \cdot \nabla \delta(\hat{n} \cdot x - b) dx \\
&= \beta \int_{\mathbb{R}^n} \hat{n} \cdot \nabla V e^{-\beta V} (\delta \hat{n} \cdot x^{\perp} - \delta b) \delta(\hat{n} \cdot x - b) dx,
\end{aligned} \tag{111}$$

where for any $z \in \mathbb{R}^n$, $z^\perp = z - (z \cdot \hat{n})\hat{n}$ is the in-plane projection of z, and we used the property that the admissible variations of $\hat{n}$ need to preserve the normalization $|\hat{n}| = 1$, i.e. they need to be perpendicular to $\hat{n}$. Collecting the terms proportional to $\delta\hat{n}$ and δb in (111) we arrive at (21).

B Derivation of (81)

It is a standard result [Pap76] that in the limit as $\delta \to 0$, (80) converges to

$$\dot{Q}_j = b_j(Q) + \sqrt{2\beta^{-1}a_j(Q)}\,\eta_j, \quad j = 1, \ldots, m. \tag{112}$$

where $a_j(q)$ is given by (83) and

$$b_j(q) = \langle -\nabla V \cdot \nabla q_j + \beta^{-1}\Delta q_j \rangle_{q(x)=q}. \tag{113}$$

Since

$$\begin{aligned}
&\int_{\mathbb{R}^n} \left(-\nabla V \nabla q_j + \beta^{-1}\Delta q_j\right) e^{-V} \delta(q_1 - q_1(x)) \cdots \delta(q_m - q_m(x)) dx \\
&\quad = \beta^{-1} \int_{\mathbb{R}^n} \nabla \cdot \left(e^{-\beta V} \nabla q_j\right) \delta(q_1 - q_1(x)) \cdots \delta(q_m - q_m(x)) dx \\
&\quad = -\beta^{-1} \int_{\mathbb{R}^n} e^{-\beta V} \nabla q_j \cdot \nabla \left(\delta(q_1 - q_1(x)) \cdots \delta(q_m - q_m(x))\right) dx \\
&\quad = \beta^{-1} \frac{\partial}{\partial q_j} \int_{\mathbb{R}^n} e^{-\beta V} |\nabla q_j|^2 \delta(q_1 - q_1(x)) \cdots \delta(q_m - q_m(x)) dx
\end{aligned} \tag{114}$$

where we used (57), one has

$$b_j(q) e^{-\beta F} = \beta^{-1} \frac{\partial}{\partial q_j} \left(a_j(q) e^{-\beta F}\right), \tag{115}$$

or, equivalently,

$$b_j(q) = -a_j(q) \frac{\partial F_j}{\partial q_j} + \beta^{-1} \frac{\partial a_j}{\partial q_j}. \tag{116}$$

Inserting this equality in (112) gives (81).

C The calculations of section 11

To show that the relaxation rate in (95) is precisely the one obtained from the accelerated dynamics equation in (81), note that since

$$\int_{\mathbb{R}^n} |\nabla q|^2 e^{-\beta V} \delta(q - q(x)) dx = \int_{q(x)=q} |\nabla q| e^{-\beta V} d\sigma(x) = cst \tag{117}$$

for the solution of (96), it follows that the relaxation rate in (95) can also be expressed as

$$\begin{aligned}\lambda_1 &= Z^{-1}\frac{N_1+N_2}{N_1N_2}\int_0^1\Big(\int_{q(x)=q}|\nabla q|e^{-\beta V}d\sigma(x)\Big)^{-1}dq \\ &= \frac{N_1+N_2}{N_1N_2}\int_0^1 a^{-1}(q)e^{\beta F(q)}dq\end{aligned} \tag{118}$$

where $a(q)$ is given by (83). (118) is precisely the asymptotic expression for the relaxation rate associated with (81) when B_1 and B_2 are metastable (see e.g. [Gar89]).

To show that the minimizer of I_1 is a maximizer of I_2, we compute the first variation of I_2 with respect to $q(x)$. Starting from the following expression equivalent to (100)

$$I_2 = \int_0^1\Big(\int_{\mathbb{R}^n}|\nabla q|^2e^{-\beta V}\delta(q(x)-q))dx\Big)^{-1}dq, \tag{119}$$

we obtain

$$\begin{aligned}\delta I_2 = &-2\int_0^1 A^{-2}(q)\int_{\mathbb{R}^n}\nabla\delta q\cdot\nabla q e^{-\beta V}\delta(q(x)-q))dxdq \\ &-\int_0^1 A^{-2}(q)\int_{\mathbb{R}^n}|\nabla q|^2e^{-\beta V}\delta q\,\delta'(q(x)-q))dxdq\end{aligned} \tag{120}$$

where

$$A(q) = \int_{\mathbb{R}^n}|\nabla q|^2e^{-\beta V}\delta(q(x)-q))dx = \int_{q(x)=q}|\nabla q|e^{-\beta V}d\sigma(x). \tag{121}$$

By integration by parts of the first integral in (120) we obtain

$$\begin{aligned}\delta I_2 = &2\int_0^1 A^{-2}(q)\int_{\mathbb{R}^n}\delta q\nabla\cdot\Big(e^{-\beta V}\nabla q\Big)\delta(q(x)-q))dxdq \\ &+\int_0^1 A^{-2}(q)\int_{\mathbb{R}^n}|\nabla q|^2e^{-\beta V}\delta q\,\delta'(q(x)-q))dxdq\end{aligned} \tag{122}$$

This can also be written as

$$\begin{aligned}\delta I_2 = &2\int_0^1 A^{-2}(q)\int_{\mathbb{R}^n}\delta q\nabla\cdot\Big(e^{-\beta V}\nabla q\Big)\delta(q(x)-q))dxdq \\ &-\int_0^1 A^{-2}(q)\frac{\partial}{\partial q}\int_{\mathbb{R}^n}|\nabla q|^2e^{-\beta V}\delta q\,\delta(q(x)-q))dxdq,\end{aligned} \tag{123}$$

which after integration by parts in q of the second integral gives

$$\begin{aligned}\delta I_2 = &2\int_0^1 A^{-2}(q)\int_{\mathbb{R}^n}\delta q\nabla\cdot\Big(e^{-\beta V}\nabla q\Big)\delta(q(x)-q))dxdq \\ &-2\int_0^1 A^{-3}(q)\frac{\partial A}{\partial q}\int_{\mathbb{R}^n}|\nabla q|^2e^{-\beta V}\delta q\,\delta(q(x)-q))dxdq.\end{aligned} \tag{124}$$

Since

$$\begin{aligned}\frac{\partial A}{\partial q} &= -\int_{\mathbb{R}^n} |\nabla q|^2 e^{-\beta V}\delta'(q(x)-q))dx \\ &= -\int_{\mathbb{R}^n} e^{-\beta V}\nabla q\cdot\nabla\delta(q(x)-q))dx \\ &= \int_{\mathbb{R}^n} \nabla\cdot\left(e^{-\beta V}\nabla q\right)\delta(q(x)-q))dx.\end{aligned} \tag{125}$$

it follows that if

$$\nabla\cdot\left(e^{-\beta V}\nabla q\right) = e^{-\beta V}|\nabla q|^2 H(q) \tag{126}$$

for an arbitrary H, then

$$\begin{aligned}\delta I_2 = 2\int_0^1 A^{-2}(q)H(q)\int_{\mathbb{R}^n}\delta q|\nabla q|^2\delta(q(x)-q))dxdq \\ -2\int_0^1 A^{-2}(q)H(q)\int_{\mathbb{R}^n}|\nabla q|^2 e^{-\beta V}\delta q\,\delta(q(x)-q))dxdq \\ = 0.\end{aligned} \tag{127}$$

Note that (126) with $H(q) = -\beta^{-1}f''(q)/f'(q)$ is precisely the equation obtained from (96) after the gauge transformation $q \to \bar{q} = f(q)$. By computing the second order variation of I_2 around the solution of (126), one can actually show that this $q(x)$ maximizes I_2.

Finally we derive (104), then (106) by maximizing (104). By assumption, $q(x) = q(\varphi(\alpha))$ provided that $\hat{n}\cdot(x-\varphi) = 0$. It follows that

$$|\nabla q(x)| = g'(\alpha)\left(\hat{n}\cdot\varphi_\alpha - \hat{n}_\alpha\cdot(x-\varphi)\right)^{-1}, \tag{128}$$

where $g(\alpha) = q(\varphi(\alpha))$. Because of (102) and (103), the term $\hat{n}_\alpha\cdot(x-\varphi)$ is a small correction to $\hat{n}\cdot\varphi_\alpha$ and we will approximate (128) by

$$|\nabla q(x)| = g'(\alpha)\left(\hat{n}\cdot\varphi_\alpha\right)^{-1}, \tag{129}$$

This is a localization approximation which is valid provided that the transition tube T defined in (98) is not too wide. Using (129) and noting that $dq = g'(\alpha)d\alpha$ if α instead of q is used as integration variable, (100) can be written as

$$\begin{aligned}I_2 &= \int_0^1\left(\int_{\mathbb{R}^n} e^{-\beta V}g'(\alpha)(\hat{n}\cdot\varphi_\alpha)^{-1}\delta(\hat{n}\cdot(x-\varphi))dx\right)^{-1}g'(\alpha)d\alpha, \\ &= \int_0^1\left(\int_{\mathbb{R}^n} e^{-\beta V}\delta(\hat{n}\cdot(x-\varphi))dx\right)^{-1}\hat{n}\cdot\varphi_\alpha d\alpha,\end{aligned} \tag{130}$$

(130) is identical to (104), and it can also be written as

$$I_2 = \int_0^1 e^{\beta F(\alpha)}\hat{n}\cdot\varphi_\alpha d\alpha \tag{131}$$

where

$$F(\alpha) = -\beta^{-1} \log \int_{\mathbb{R}^n} e^{-\beta V} \delta(\hat{n} \cdot (x - \varphi)) dx \tag{132}$$

We must maximizes (130) with respect to $\hat{n}$ subject to the constraint (102). This constraint is explicitly

$$0 = \int_{\mathbb{R}^n} (x - \varphi) e^{-\beta V(x)} \delta(\hat{n} \cdot (x - \varphi)) dx, \tag{133}$$

and we enforce it by means of adding a Lagrange multiplier term in (130):

$$\begin{aligned} I_2 = &\int_0^1 \Big(\int_{\mathbb{R}^n} e^{-\beta V} \delta(\hat{n} \cdot (x - \varphi)) dx \Big)^{-1} \hat{n} \cdot \varphi_\alpha d\alpha \\ &+ \int_0^1 \int_{\mathbb{R}^n} \lambda \cdot (x - \varphi) e^{-\beta V(x)} \delta(\hat{n} \cdot (x - \varphi)) dx d\alpha. \end{aligned} \tag{134}$$

where $\lambda = \lambda(\alpha)$ is the Lagrange multiplier to be determined later. To obtain the Euler-Lagrange equations associated with (134), we compute the first variation of this functional with respect to $\hat{n}$ and φ. Since $|\hat{n}| = 1$, we can restrict ourselves variations of the type $\hat{n} \to \hat{n} + \delta \hat{n}^\perp$ in $\hat{n}$, which gives

$$
\begin{aligned}
\delta_{\hat{n}} I_2 &= \int_0^1 e^{\beta F} \delta\hat{n}^\perp \cdot \varphi_\alpha d\alpha \\
&\quad - \int_0^1 e^{2\beta F} \Big(\int_{\mathbb{R}^n} e^{-\beta V} \delta\hat{n}^\perp \cdot (x-\varphi) \delta'(\hat{n} \cdot (x-\varphi)) dx \Big) \hat{n} \cdot \varphi_\alpha d\alpha \\
&\quad + \int_0^1 \int_{\mathbb{R}^n} \lambda \cdot (x-\varphi) e^{-\beta V(x)} \delta\hat{n}^\perp \cdot (x-\varphi) \delta'(\hat{n} \cdot (x-\varphi)) dx d\alpha \\
&= \int_0^1 e^{\beta F} \delta\hat{n}^\perp \cdot \varphi_\alpha d\alpha \\
&\quad - \int_0^1 e^{2\beta F} \Big(\int_{\mathbb{R}^n} e^{-\beta V} \delta\hat{n}^\perp \cdot (x-\varphi) \hat{n} \cdot \nabla \delta(\hat{n} \cdot (x-\varphi)) dx \Big) \hat{n} \cdot \varphi_\alpha d\alpha \\
&\quad + \int_0^1 \int_{\mathbb{R}^n} \lambda \cdot (x-\varphi) e^{-\beta V(x)} \delta\hat{n}^\perp \cdot (x-\varphi) \hat{n} \cdot \nabla \delta(\hat{n} \cdot (x-\varphi)) dx d\alpha \\
&= \int_0^1 e^{\beta F} \delta\hat{n}^\perp \cdot \varphi_\alpha d\alpha \\
&\quad - \beta \int_0^1 e^{2\beta F} \Big(\int_{\mathbb{R}^n} e^{-\beta V} (\hat{n} \cdot \nabla V) \delta\hat{n}^\perp \cdot (x-\varphi) \delta(\hat{n} \cdot (x-\varphi)) dx \Big) \hat{n} \cdot \varphi_\alpha d\alpha \\
&\quad + \beta \int_0^1 \int_{\mathbb{R}^n} \lambda \cdot (x-\varphi) e^{-\beta V(x)} (\hat{n} \cdot \nabla V) \delta\hat{n}^\perp \cdot (x-\varphi) \delta(\hat{n} \cdot (x-\varphi)) dx d\alpha \\
&= \int_0^1 e^{\beta F} \delta\hat{n}^\perp \cdot \varphi_\alpha d\alpha \\
&\quad - \beta \int_0^1 e^{\beta F} \left\langle (\hat{n} \cdot \nabla V) \delta\hat{n}^\perp \cdot (x-\varphi) \right\rangle_{\bar{P}(\alpha)} \hat{n} \cdot \varphi_\alpha d\alpha \\
&\quad + \beta \int_0^1 e^{-\beta F} \left\langle \lambda \cdot (x-\varphi) (\hat{n} \cdot \nabla V) \delta\hat{n}^\perp \cdot (x-\varphi) \right\rangle_{\bar{P}(\alpha)} d\alpha
\end{aligned}
\tag{135}
$$

where we used (132) and (133) repeatedly. It follows that the maximizer of (134) satisfies

$$
\begin{aligned}
(\varphi_\alpha)^\perp &= \beta \left\langle (x-\varphi)(\hat{n} \cdot \nabla V) \right\rangle_{\bar{P}(\alpha)} \hat{n} \cdot \varphi_\alpha \\
&\quad - \beta e^{-2\beta F} \left\langle (x-\varphi) \lambda \cdot (x-\varphi) (\hat{n} \cdot \nabla V) \right\rangle_{\bar{P}(\alpha)}.
\end{aligned}
\tag{136}
$$

Similarly, by computing the variation of (134) with respect to φ and restricting ourselves to variations of the type $\hat{\varphi} \to \hat{n} + \delta\hat{\varphi}^\perp$ (since we are looking to the new position of $\varphi(\alpha)$ in $\bar{P}(\alpha)$ and therefore $\varphi(\alpha)$ needs to stay in this plane by definition) we obtain

$$
\begin{aligned}
\delta_\varphi I_2 &= \int_0^1 e^{\beta F} \hat{n} \cdot (\delta\varphi^\perp)_\alpha d\alpha \\
&\quad - \int_0^1 \int_{\mathbb{R}^n} \lambda(\alpha) \cdot \delta\varphi^\perp e^{-\beta V(x)} \delta(\hat{n} \cdot (x-\varphi)) dx d\alpha \\
&= - \int_0^1 e^{\beta F} \hat{n}_\alpha \cdot \delta\varphi^\perp d\alpha - \int_0^1 \lambda^\perp \cdot \delta\varphi e^{-\beta F} d\alpha.
\end{aligned}
\tag{137}
$$

Therefore the maximizer of (134) must also satisfy

$$0 = e^{\beta F}\hat{n}_\alpha + \lambda^\perp e^{-\beta F} \tag{138}$$

Solving this equation in $\lambda^\perp$ and inserting the result in (136), we arrive at

$$\begin{aligned}(\varphi_\alpha)^\perp = {} & \beta \left\langle (x-\varphi)(\hat{n}\cdot\nabla V)\right\rangle_{\bar{P}(\alpha)} \hat{n}\cdot\varphi_\alpha \\ & + \beta \left\langle (x-\varphi)\hat{n}_\alpha\cdot(x-\varphi)(\hat{n}\cdot\nabla V)\right\rangle_{\bar{P}(\alpha)}.\end{aligned} \tag{139}$$

Combined with (102) this equation specifies $\hat{n}$ and φ. Next we show that (139) is strictly equivalent to (105). First notice that

$$\left\langle (x-\varphi)(\hat{n}\cdot\nabla V)\hat{n}_\alpha\cdot(x-\varphi)\right\rangle_{\bar{P}(\alpha)} = \left\langle (x-\varphi)(\varphi_\alpha\cdot\nabla V)\right\rangle_{\bar{P}(\alpha)}. \tag{140}$$

Indeed

$$\begin{aligned}&\beta\langle (x-\varphi)(\hat{n}\cdot\nabla V)\hat{n}_\alpha\cdot(x-\varphi)\rangle_{\bar{P}(\alpha)} \\ &\quad = \beta e^{-\beta F}\int_{\mathbb{R}^n}(x-\varphi)\hat{n}_\alpha\cdot(x-\varphi)(\hat{n}\cdot\nabla V)e^{-\beta V}\delta(\hat{n}\cdot(x-\varphi))dx \\ &\quad = -e^{-\beta F}\int_{\mathbb{R}^n}\hat{n}\cdot\nabla\Big((x-\varphi)\hat{n}_\alpha\cdot(x-\varphi)e^{-\beta V}\Big)\delta(\hat{n}\cdot(x-\varphi))dx \\ &\quad = e^{-\beta F}\int_{\mathbb{R}^n}(x-\varphi)\hat{n}_\alpha\cdot(x-\varphi)e^{-\beta V}\delta'(\hat{n}\cdot(x-\varphi))dx\end{aligned} \tag{141}$$

where we used (133) to get the second equality. Since

$$\begin{aligned}&\hat{n}_\alpha\cdot(x-\varphi)\delta'(\hat{n}\cdot(x-\varphi)) \\ &\quad = \frac{\partial}{\partial\alpha}\delta(\hat{n}\cdot(x-\varphi)) + \hat{n}\cdot\varphi_\alpha\delta'(\hat{n}\cdot(x-\varphi)) \\ &\quad = \frac{\partial}{\partial\alpha}\delta(\hat{n}\cdot(x-\varphi)) + \varphi_\alpha\cdot\nabla\delta(\hat{n}\cdot(x-\varphi)),\end{aligned} \tag{142}$$

one can continue as

$$\begin{aligned}&\beta\langle (x-\varphi)(\hat{n}\cdot\nabla V)\hat{n}_\alpha\cdot(x-\varphi)\rangle_{\bar{P}(\alpha)} \\ &\quad = e^{-\beta F}\int_{\mathbb{R}^n}(x-\varphi)e^{-\beta V}\frac{\partial}{\partial\alpha}\delta(\hat{n}\cdot(x-\varphi))dx \\ &\qquad + e^{-\beta F}\int_{\mathbb{R}^n}(x-\varphi)e^{-\beta V}\varphi_\alpha\cdot\nabla\delta(\hat{n}\cdot(x-\varphi))dx \\ &\quad = e^{-\beta F}\frac{\partial}{\partial\alpha}\int_{\mathbb{R}^n}(x-\varphi)e^{-\beta V}\delta(\hat{n}\cdot(x-\varphi))dx + \varphi_\alpha \\ &\qquad + e^{-\beta F}\int_{\mathbb{R}^n}(x-\varphi)e^{-\beta V}\varphi_\alpha\cdot\nabla\delta(\hat{n}\cdot(x-\varphi))dx \\ &\quad = \beta e^{-\beta F}\int_{\mathbb{R}^n}(x-\varphi)(\varphi_\alpha\cdot\nabla V)e^{-\beta V}\delta(\hat{n}\cdot(x-\varphi))dx \\ &\quad = \beta\left\langle (x-\varphi)(\varphi_\alpha\cdot\nabla V)\right\rangle_{\bar{P}(\alpha)}\end{aligned} \tag{143}$$

and arrive at (140). Second note that

$$(\varphi_\alpha)^\perp = \beta\langle((\varphi_\alpha)^\perp \cdot \nabla V)(x-\varphi)\rangle. \tag{144}$$

To see this we start from

$$J = e^{-\beta F}\int_{\mathbb{R}^n} e^{-V}(x-\varphi)(\hat{n}\cdot\varphi_\alpha)\delta'(\hat{n}\cdot(x-\varphi))dx \tag{145}$$

and use either

$$(\hat{n}\cdot\varphi_\alpha)\delta'(\hat{n}\cdot(x-\varphi)) = (\hat{n}\cdot\varphi_\alpha)\hat{n}\cdot\nabla\delta'(\hat{n}\cdot(x-\varphi)) \tag{146}$$

or

$$(\hat{n}\cdot\varphi_\alpha)\delta'(\hat{n}\cdot(x-\varphi)) = \varphi_\alpha\cdot\nabla\delta'(\hat{n}\cdot(x-\varphi)) \tag{147}$$

together with (146) gives

$$\begin{aligned} J &= e^{\beta F}\int_{\mathbb{R}^n} e^{-\beta V}(x-\varphi)(\hat{n}\cdot\varphi_\alpha)\hat{n}\cdot\nabla\delta(\hat{n}\cdot(x-\varphi)) \\ &= -\hat{n}(\hat{n}\cdot\varphi_\alpha) + \beta\langle(x-\varphi)(\hat{n}\cdot\nabla V)\rangle_{\bar{P}(\alpha)}(\hat{n}\cdot\varphi_\alpha) \end{aligned} \tag{148}$$

If we use (147) instead, we get

$$\begin{aligned} J &= e^{\beta F}\int_{\mathbb{R}^n} e^{-\beta V}(x-\varphi)(\hat{n}\cdot\varphi_\alpha)\hat{n}\cdot\nabla\delta(\hat{n}\cdot(x-\varphi)) \\ &= \varphi_\alpha + \beta\langle(x-\varphi)(\varphi_\alpha\cdot\nabla V)\rangle_{\bar{P}(\alpha)} \end{aligned} \tag{149}$$

(148) and (149) gives (144). Inserting (140) and (144) in (139) we deduce that this equation is indeed equivalent to (105).

Finally to derive (106), we recall that (130) is equivalent to the original object function in (100) only within the localization assumption in (103) which allows us to approximate (128) by (129). But (103) implies that

$$\frac{\partial}{\partial\alpha}\delta(\hat{n}\cdot(x-\varphi)) \approx -\hat{n}\cdot\varphi_\alpha\delta'(\hat{n}\cdot(x-\varphi)), \tag{150}$$

since the term $\hat{n}_\alpha\cdot(x-\varphi)\delta'(\hat{n}\cdot(x-\varphi))$ in (142) is a small correction. Using (150), we have

$$
\begin{aligned}
0 &= \beta \langle (\hat{n}\cdot\nabla V)(x-\varphi)\rangle_{\bar{P}(\alpha)} \\
&= \beta e^{-\beta F}\int_{\mathbb{R}^n} (\hat{n}\cdot\nabla V) e^{-\beta V}(x-\varphi)\delta(\hat{n}\cdot(x-\varphi))dx \\
&= \hat{n} - e^{-\beta F}\int_{\mathbb{R}^n} \hat{n}\cdot\nabla\Big(e^{-\beta V}(x-\varphi)\Big)\delta(\hat{n}\cdot(x-\varphi))dx \\
&= \hat{n} + e^{-\beta F}\int_{\mathbb{R}^n} e^{-\beta V}(x-\varphi)\delta'(\hat{n}\cdot(x-\varphi))dx \\
&= \hat{n} - e^{-\beta F}(\hat{n}\cdot\varphi_\alpha)^{-1}\int_{\mathbb{R}^n} e^{-\beta V}(x-\varphi)\frac{\partial}{\partial\alpha}\delta(\hat{n}\cdot(x-\varphi))dx \\
&= \hat{n} - e^{-\beta F}(\hat{n}\cdot\varphi_\alpha)^{-1}\frac{\partial}{\partial\alpha}\int_{\mathbb{R}^n} e^{-\beta V}(x-\varphi)\delta(\hat{n}\cdot(x-\varphi))dx - (\hat{n}\cdot\varphi_\alpha)^{-1}\varphi_\alpha \\
&= \hat{n} - (\hat{n}\cdot\varphi_\alpha)^{-1}\varphi_\alpha .
\end{aligned}
\tag{151}
$$

This is (106).

References

[Ben77] C. H. Bennett, "Exact defect calculations in model substances" in A. S. Nowick and J. J. Burton, editors, *Algorithms for Chemical Computation*, ACS Symposium Series No. 46, **63** (1977).

[BEGK03a] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein, "Metastability in reversible diffusion processes 1. Sharp estimates for capacities and exit times," WIAS-preprint 767 (2003).

[BEGK03b] A. Bovier, V. Gayrard, and M. Klein, "Metastability in reversible diffusion processes. 2. Precise estimates for small eigenvalues," WIAS-preprint 768 (2003).

[BCDG02] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, "Transition Path Sampling: Throwing Ropes over Mountain Passes in the Dark," *Annual Review in Physical Chemistry* **53**, 291 (2002).

[BDC98] P. G. Bolhuis, C. Dellago and D. Chandler, "Sampling ensembles of deterministic transition pathways," *Faraday Discussion* **110**, 421 (1998).

[BDC00] P. G. Bolhuis, C. Dellago, and D. Chandler, "Reaction Coordinates of Biomolecular Isomerization," *Proc. of Nat. Acad. Sci. USA* **97**, 5877 (2000).

[CCHK89] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral' " Constrained reaction coordinate dynamics for the simulation of rare events," *Chem. Phys. Lett.* **156**, 472 (1989).

[Chan78] D. Chandler, *J. Chem. Phys.* **68**, 2959 (1978).

[DBG02] C. Dellago, P. G. Bolhuis, P. L. Geissler, "Transition Path Sampling," *Advances in Chemical Physics* **123**, 1 (2002).

[ERVE02a] W. E, W. Ren, and E. Vanden-Eijnden, "String method for the study of rare events," *Phys. Rev. B* **66**, 052301 (2002).

[ERVE02b] W. E, W. Ren, and E. Vanden-Eijnden, "Energy Landscapes and Rare Events," ICM report (2002).

[EGC02] R. Elber, A. Ghosh, and A. Cárdenas, "Long time dynamics of complex systems", *Account of Chemical Research* **35**,396–403 (2002).

[ECGS03] R. Elber, A. Ghosh, A. Cárdenas, and H. Stern, "Bridging the gap between reaction pathways, long time dynamics and calculation of rates", *Advances in Chemical Physics* **126**, 93–129 (2003).

[Eyr35] H. Eyring, *J. Chem. Phys.* , **3** 107 (1935).

[FSDC02] A. Fischer, Ch. Schütte, P. Deuflhard, and F. Cordes, "Hierarchical Uncoupling-Coupling of Metastable Conformations Computational Methods for Macromolecules" in: *Challenges and Applications*, Proceedings of the 3rd International Workshop on Algorithms for Macromolecular Modeling. Edited by T. Schlick and H. H. Gan, LNCSE, Springer, 24 (2002)

[FW98] M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems*, 2nd ed. (Springer, 1998).

[FS01] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithm to Applications*, 2nd. ed. (Elsevier, 2001).

[Gar89] G. W. Gardiner, *Handbook of stochastic methods*, 2nd ed. (Springer, 1997).

[HJ00] G. Henkelman and H. Jónsson, "Improved Tangent Estimate in the NEB Method for Finding Minimum Energy Paths and Saddle Points," *J. Chem. Phys.* **113**, 9978 (2000).

[HTV94] T. Huber, A. E. Torda, and W.F. van Gunsteren, "Local elevation: A method for improving the searching properties of molecular dynamics simulation," *J. Comput.-Aided Mol. Design* **8**, 695 (1994).

[HMS02] W. Huisinga, S. Meyn, and Ch. Schütte, "Phase Transitions and Metastability in Markovian and Molecular Systems," submitted (2002).

[ILP03] M. Iannuzzi, A. Laio, and M. Parrinello, "Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics," *Phys. Rev. Lett.* **90**, 238302 (2003)

[JJ01] G. H. Jóhannesson and H. Jónsson, "Optimization of Hyperplanar Transition States," *J. Chem. Phys.* **115**, 9644 (2001).

[JMJ98] H. Jónsson, G. Mills, K. W. Jacobsen, "Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions," p. 385 in *Classical and Quantum Dynamics in Condensed Phase Simulations*, ed. B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, 1998)

[LP02] A. Laio and M. Parrinello, "Escaping free-energy minima," *Proc. Nat. Acad. Sc. USA* **99**, 12562 (2002).

[Pap76] G. Papanicolaou, "Some probabilistic problems and methods in singular perturbations," *Rocky Mountain Math. J.* **6**, 653 (1976).

[Pra86] L. R. Pratt, "A statistical method for identifying transition states in high dimensional problems," *J. Chem. Phys.* **9**, 5045 (1986).

[Ren02] W. Ren, *Numerical Methods for the Study of Energy Landscapes and Rare Events*, PhD thesis, New York University, 2002.

[Ren03] W. Ren, "Higher order string method for finding minimum energy paths," *Comm. Math. Sci.* **1**, 377 (2003).

[RCB77] J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, "Numerical integration of the Cartesian equation of motion of a system with constraints: molecular dynamics of N-alkanes," *J. Comp. Phys.* **23**, 327 (1977).

[SH02] Ch. Schütte and W. Huisinga, "Biomolecular Conformations can be Identified as Metastable Sets of Molecular Dynamics," submitted (2002).

[SHD02] Ch. Schütte, W. Huisinga, and P. Deuflhard, "Transfer Operator Approach to Conformational Dynamics in Biomolecular Systems," pp.191–223 in *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems* edited by B. Fiedler (2001)

[SV00] M. R. Sorensen and A. Voter, "Temperature-accelerated dynamics for simulation of infrequent events," *J. Chem. Phys.* **112**, 9599 (2000).

[SC98] M. Sprik and G. Ciccotti, "Free energy from constrained molecular dynamics," *J. Chem Phys.* **109**, 7737 (1998).

[Str01] J. E. Straub, "Reaction rates nd transition pathways", p. 199 in *Computational biochemistry and biophysics*, ed. O. M Becker, A. D MacKerell, Jr., B. Roux, and M. Watanabe (Marcel Decker, Inc. 2001).

[UE89] A. Ulitsky and R. Elber, "A new technique to calculate steepest descent paths in flexible polyatomic systems," *J. Chem. Phys.* **92**, 1510 (1990).

[Vot97] A. Voter, "Hyperdynamics: Accelerated molecular dynamics of infrequent events," *Phys. Rev. Lett.* **78**, 3908 (1997).

[WL01] F. Wang and P. D. Landau, "Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States," *Phys. Rev. Lett.* **86**, 2050 (2001).

[WCH01] T. O. White, G. Ciccotti, and J.-P. Hansen, "Brownian dynamics with constraints." *Mol. Phys.* **99**, 2023, (2001)

[Wig38] E. Wigner, *Trans. Faraday Soc.* **34**, 29 (1938).

[Yam60] T. Yamamoto, *J. Chem Phys.* , **33**, 281 (1960).

Nonlinear Dynamics Analysis through Molecular Dynamics Simulations

Ioannis G. Kevrekidis[1], Ju Li[2], and Sidney Yip[3]

[1] Department of Chemical Engineering, PACM and Mathematics, Princeton University, Princeton, New Jersey 08544, USA yannis@princeton.edu
[2] Department of Materials Science and Engineering, Ohio State University, Columbus, Ohio 43210, USA li.562@osu.edu
[3] Department of Nuclear Engineering, Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA syip@mit.edu

A general goal in multiscale modeling is to analyze macroscale (system-level) phenomena using information on the system at the microscale. In this pedagogical note we revisit the time-stepper approach to performing nonlinear dynamics analysis through the use of a microscopic simulator, molecular dynamics in this case. Using simple illustrative examples we suggest that there are better ways of using a molecular simulator than observing temporal evolution of the system in a hands-off manner alone. Continuum numerical analysis algorithms can be transformed into alternative computational protocols for microscopic solvers: macroscopic nonlinear dynamics information, such as stationary solutions, stability boundaries or similarity exponents, can be obtained through the design and execution of appropriately initialized short bursts of direct microscopic simulation.

1 Introduction

In the computer-assisted study of nonlinear dynamical systems, direct temporal simulation is not the only available approach. A host of computational tools, such as numerical bifurcation analysis, aimed at efficiently extracting quantitative information have been developed. Such tools include Newton-Raphson and other fixed point algorithms that accelerate the location of steady states, continuation of solution branches in parameter space, eigensolvers that quantify stability, boundary value solvers to accelerate the location of limit cycles, algorithms for codimension-one or higher order bifurcation points to locate transitions, etc. When explicit evolution equations ("macro"-level equations for the purpose of this note) are available, such techniques (given a good initial guess) can be remarkably efficient compared to direct temporal simulation. Over the last ten years a trend has arisen in continuum numerical analysis towards so-called *time-stepper based* methods, like the Recursive Projection Method (RPM) introduced by Shroff and Keller [SK93], see also [TB00]. These

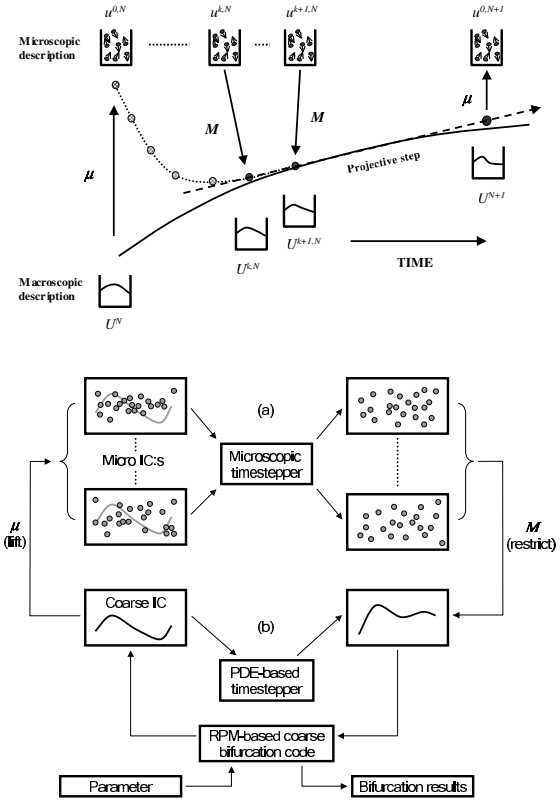
T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

methods are aimed at enabling deterministic dynamic simulators to perform tasks beyond direct temporal simulation: bifurcation analysis, but also long-term prediction, stability analysis, control and optimization. This is achieved through computational protocols implemented through a software superstructure: what, in this discussion, we will call "the wrapper". The wrapper is then a way of combining continuum numerical tools, such as bifurcation analysis techniques, as the *outer* component, with a time integrator (time stepper), as the *inner* component. The functional relation of the two components is indicated schematically in Fig. 1 for the case of coarse projective integration and coarse bifurcation analysis [TQK00, GKT02].

Fig. 1 goes beyond the *deterministic* time-stepper case: it illustrates the use of this wrapper-based computational enabling technology. System level coarse "outer" algorithms (integration, RPM or GMRES-based fixed point location) are wrapped around inner fine scale *atomistic / stochastic* timesteppers. It is interesting that the same wrapper can be operated transparently in two modes, depending on whether a deterministic coarse equation is available or not. For a given set of parameters the time-stepper component is called along with a coarse initial condition, which is evolved forward in time -with a deterministic coarse PDE solver- to produce a coarse output (see Fig. 1b). The results are processed, and new trial initial conditions are constructed and evolved; after this procedure is repeated a sufficient number of times, bifurcation results are obtained. When a coarse equation is not available the time-stepper protocol operates transparently, but with one additional feature - the "lifting" of the coarse initial condition to one or more consistent microscale initial conditions, and the "restriction" of evolved microscale information back to macroscopic observables.

The lifting step is shown on the left in both parts of Fig. 1. For each microscale initial condition thus produced, the system is evolved forward using the microscopic timestepper, and the evolved microscale conditions are "restricted" to give a coarse output that is then returned to the integration or bifurcation code. Notice that the first mode requires the availability of coarse PDE, whereas such equations are not needed in the second mode: the coarse input-output map for the macroscopic observable has been *estimated* through short computational experimentation with the micro solver. The microscopic timestepper mode implicitly takes into account microscale processes in the bifurcation analysis; it makes the analysis feasible even if the appropriate coarse PDE for the system is not known. On the other hand, it remains to be shown that good results can be obtained using the microscopic timestepper. The distinction between coarse (PDE-based) and detailed (microscopic) timesteppers has been discussed in [TQK00, GKT02, RTK02, MMPK02] and reviewed in [KGHKRT03].

We are interested in the use of molecular dynamics (MD) simulation, a well-known method for following the dynamical evolution of a system of atoms and molecules, as the inner, microscopic timestepper. Since molecular dynamics is not designed to perform system-level analysis, running the MD simulation in the conventional manner does not lead naturally to results on nonlinear dynamics behavior on the coarse scale. In what follows we argue that a wrapper can be constructed to enable bifurcation analysis to be performed with MD as the timestepper. We use



(b)

Fig. 1. Schematic of timestepper-based numerical computation: (a) coarse integration and (b) RPM-based coarse bifurcation analysis. Macro-initial conditions are *lifted* to consistent microscopic ones, evolved through the timestepper, and *restricted* back to the macroscale. The wrapper is templated on continuum numerical analysis algorithms. It designs and executes these short bursts of computational experimentation with the micro-solver, processes their results, and uses them to estimate the quantities required in the macro-numerics.

MD results from existing simulations to illustrate the general utility of the timestepper/wrapper notions, pointing to what has been done (without such notions) to show what more can be done with such notions. While the correspondence and language may not yet be as precise as one would like, it appears that the new connections being made between NLD and MD can be potentially useful.

2 Linking Scales (Practical Determinism)

What is the precise meaning of the statement that "macroscopic equations close at a certain level of description"? Colloquially this implies that one can be practically predictive at that level: given the value of "a few" system observables at a moment in time, one can predict the evolution of these observables at a later time without additional information. If equations for the evolution of these observables are not known, can we put practical determinism (i.e., the information that equations *conceptually* exist) to good use in microscopic simulations ? Our answer is that the timestepper-wrapper method does not require *explicit formulas* for the equations. There are reasons to believe that a time-stepper approach, wrapped around a microscopic simulator, can succeed in solving the equations without writing them down.

The challenge, put in a different way, lies in the linking of micro and macro scales; one would like to make macroscopic predictions on the basis of "just enough" input from the microscale. For our example we consider the prediction of the onset of elastic instability in nanoindentation of a thin-film sample, which leads to the homogeneous nucleation of a dislocation loop, a microcrack or a deformation twin [LVZYS02, VLZYS03, ZLVOYS04]. At the macro scale, the strain induced in the sample may be described in the finite-element method (FEM) by a mesh of grid points (nodes). The FEM governing equations may not be closed because the constitutive (stress-strain) relations capable of describing large-strain deformation at the nodes are generally not known in closed form. On the other hand, MD is a method of simulating the phenomenon at the micro scale without the use of equations describing the manifestation of the instability at the level of macroscopic observables. The challenge is then in passing just enough microscopic information (obtained by running the molecular simulator as parsimoniously as possible in space and time) up to the macro scale, where the manifestation of this instability can be efficiently pinpointed and practically analyzed. If we think of direct dynamic simulation as "nature", we are in a sense attempting to outsmart nature: obtain the information of interest by doing as little microscopic simulation as possible; hopefully with much less effort than direct simulation of the full system in space, time and parameter space.

Timestepping can thus form the basis of a robust procedure for scale-linking: passing information between the micro- and macroscales. Scale linking is generic to practically all fundamental studies of complex systems behavior. We mention in passing that, while in some problems the relevant macroscopic observables are obvious (e.g. concentration for chemical reactions, or stress and strain in our case), the selection of the right observables (order parameters, phase field variables, reaction coordinates) constitutes an important and difficult part of any procedure for linking scales. This selection determines the level at which we can be practically deterministic.

Our example of linking scales in the modeling of nanoindentation of metal thin film involves two distinctly different simulation methods, finite-element method (FEM) and molecular dynamics (MD) [LVZYS02, VLZYS03, ZLVOYS04]. The instability in question occurs when the indenter force reaches a critical threshold, or equivalently, when the indenter penetration reaches a critical distance. The resulting

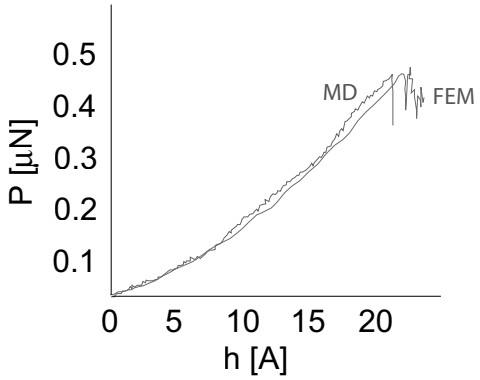


Fig. 2. Nanoindentation displacement-load (h-P) response showing initial elastic deformation of the thin film up to a critical indentation distance where the load suddenly drops (signaling the nucleation of a dislocation loop), prediction by FEM with Cauchy-Born hypothesis (smooth curve) and direct MD simulation (wiggly curve). (See Plate 7 on Page 271)

phenomenon is a sudden transition from elastic to plastic response of the thin film. To map the nanoindentation study onto the present discussion we regard FEM and MD as the coarse and detailed timesteppers respectively. In FEM the strain at each node is evolved subject to a constitutive relation that specifies the stress field. Since this information is usually not known for strains large enough to induce elastic instability, one can say that determinism does exist but the constitutive relations required to close at FEM level are not available. However, in this case it is possible to link FEM with MD using an approach known as the Cauchy-Born hypothesis. In this hypothesis one (a) converts the strain at a given node into a deformation tensor, (b) deforms a perfect crystal according to this tensor, and (c) uses an appropriate interatomic potential to calculate the stress response of the deformed infinite lattice. The resulting stress is then returned to the given node as if it were given by a constitutive relation. By applying this procedure to every node in the FEM calculation and doing it on-the-fly at every step in the evolution one arrives at a prediction of the nanoindentation response. Relative to Fig. 1 we see that the Cauchy-Born hypothesis, steps (a) through (c), are, in effect, the "lifting", "evolving" and "restriction". In Fig. 2 we show a comparison of the predicted response as simulated by the modified FEM, which we denote as interatomic potential FEM, with that obtained by direct MD simulation (detailed timestepper evolution) without projection to the coarse level. This example of linking atomistic (MD) and mesoscale (FEM) illustrates how prediction at the coarse scale can be achieved using the notion of a wrapper. Our use of the Cauchy-Born hypothesis is a special case of a general formulation linking atomistic simulation and FEM, known as the Quasicontinuum Method [TOP96, MT02]. While in this case we evaluate the atomic-level stress immediately upon lifting, in general one may want to evolve the detailed timestepper for a short time, so that errors introduced by the lifting process can "heal" [GKT02].

3 Initializing at Will

In the spirit of linking scales, we now take up the question of how to find the inflection-point and saddle-point configurations of a dynamical system. The premise here is that a saddle point is best located through fixed point schemes (like Newton-Raphson) that exploit macroscopic smoothness through the use of derivatives. Interestingly, under appropriate assumptions on separation between fast attracting and slow repelling time scales, saddles can be located by performing integration of an explicitly available equation *backward in time*. This opens up the intriguing possibility (used in MD studies of alanine dipeptide folding [HK03]) of integrating an equation backward in time to find an unstable solution [GK04]. This can be effected by initializing consistently with a macroscopic observables, evolve microscopically for a short time interval, estimate the time derivative of the macroscopic observables, and use it to project the state of the system for a longer interval *backward in time*. The overall procedure consists of short forward runs, postprocessing of their results, and reinitializations of the system at "effectively earlier in time" values of the macroscopic trajectory. Such a backward integration procedure has been used to escape free-energy minima and explore free energy surfaces; we believe it has potential in detecting transition states and exploring nucleation phenomena.

Our example for illustrating the advantages of "detailed" system initialization is the determination of the stress required to move an existing, isolated dislocation in a crystal, the so-called Peierls stress in the problem of dislocation dynamics. Since simulations generally can be initialized at will, we take advantage of this flexibility to prepare the crystal model containing a dislocation such that its subsequent behavior provides physical insight. The basic notion of preconditioning the system is not new, yet there are not many simulations that exercise this degree of control in the manner described below. Our simulation consists of setting up two runs, each involving a single dislocation in a lattice with a different sample microstructure environments (denoted as A and B). We subject both systems to incremental values of pure shear to determine the threshold shear stress at which the dislocation will start to move. Given that the two sample environments are quite distinct, it is no surprise that the threshold strain values are different for the two cases. In each run we also calculate the virtual energy release ΔE of the defect by taking a small, localized region around the defect (with dimension about 1 nm) in a manner similar to the J-integral but using atomic sums, and evaluating the energy change within this region plus the border contributions that represent the virtual work done by the environment [Li00]. Fig. 3 shows a plot of the computed local virtual energy releases ΔE with applied shear stress. It is interesting that for the two cases the critical ΔE's for A and B turn out to have the same value. This signifies that even though the sample environments A and B are different, their influences on the local environment of the dislocation is accurately portrayed through the evaluation of ΔE. ¿From the critical value of the virtual energy release, one backs out a value for the Peierls stress, in agreement with results obtained by an entirely different method [Li00, CBCLY01, CBCLY03].

This example illustrates the notion that initializing a microscopic simulation at special initial states (which is effectively impossible experimentally, but eminently

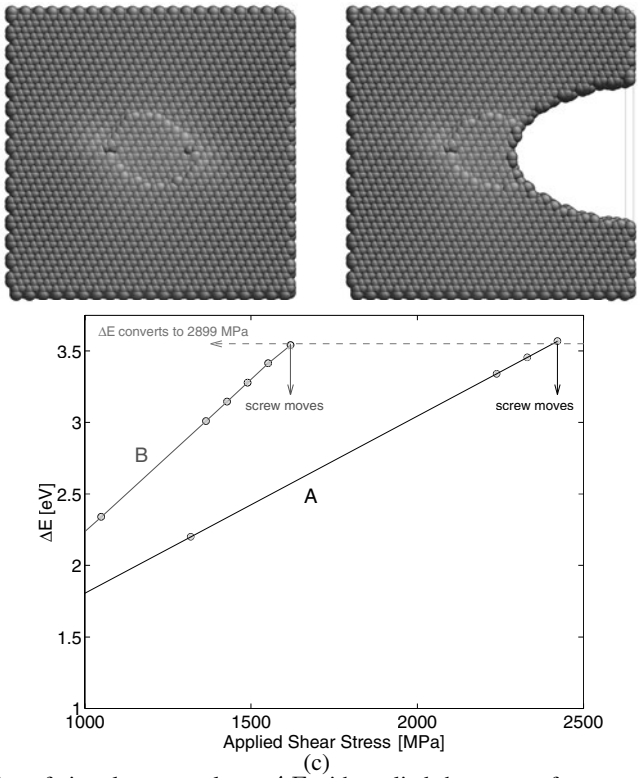


Fig. 3. Variation of virtual energy release ΔE with applied shear stress for a screw dislocation in silicon in two different microstructure environments (A and B), one in the form of a dislocation dipole, two dislocation cores as shown in (a), and the other in the form of a dislocation core with a notch nearby, as shown in (b). The left dislocation core is seen to to move at the same value of ΔE but different values of the applied shear stress. This signifies ΔE is not affected by the samepl microstructure environment and therefore it can be converted to a value for the Peierls stress that is an intrinsic material property. (See Plate 8 on Page 272)

doable computationally), evolving for short times and postprocessing the results, can be instrumental in extracting intrinsic properties of the system. By virtue of its invariance to the sample microstructure environment, the procedure demonstrates that the resulting critical energy release ΔE is indeed an intrinsic property of the material. This conclusion could not be reached in a convincing fashion by simply running the simulation of a dislocation dipole in a crystal lattice and determining the apparent critical (supercell-averaged Virial) shear stress at which the dislocation begins to move, due to image dislocation interactions [CBCLY01, CBCLY03]. Our computation of the Peierls stress did not require any knowledge of the coarse theory, i.e. equations of anisotropic elasticity; but it gave the correct coarse result based entirely on numerical evaluations of atomic sums, with a ring size as small as 1 nm. Knowing a functional of the local dislocation environment that correlates with the

onset of dislocation motion could form the basis for the systematic exploration of this onset [HSK04].

4 Dynamic rescaling and the formation of singularities.

A third issue we will examine is the question of determining self-similarity and scaling exponents from the system behavior at the onset of explosive instabilities. We believe that it is reasonable to look for rescaling space, time and observables in a way that will transform an apparently violently exploding instability scenario into a smooth stationary one. For our example we have in mind the visualization of a wave instability physically corresponding to the localization of lattice strain, as manifested in the simulation of the homogeneous nucleation of a dislocation loop or deformation twin embryo. We argue that coarse timestepping and successive rescaling (dynamic coarse renormalization) can transform the increasingly steepening of the wave front (which apparently explodes in finite time) to a time-invariant profile [ABK04, SKK03, RM00, RKML03, CBGK03].

Just as in the first example, where the system undergoes an elastic to plastic response transition in the form of dislocation nucleation under indentation, we may also ask for details of the transition from affine shear deformation to the nucleation of a deformation twin, say in the $\{\bar{1}\bar{1}2\}\langle 111\rangle$ shear system of BCC Mo [Chang03]. In contrast to dislocation nucleation which involves the relative shear of two adjacent planes, twinning is a competing process which involves relative shear among three or more adjacent planes. Recent MD simulations have shown that twinning can be homogeneously nucleated by shearing a single crystal [Chang03]. By treating the relative shear between two adjacent planes as a reaction coordinate one can formulate a one-dimensional model (a chain of these coordinates) to visualize the onset of twinning as a dynamical process of strain localization. Fig. 4 shows MD results on dislocation loop / deformation twin nucleation through the profile evolution of a small sinusoidal wave superimposed along the reaction coordinates. What one sees is a four-stage scenario: linear wave, nonlinear wave which begins to steepen, strongly singular behavior just prior to shock wave formation, and emergence of a localized shear. We believe that dynamic renormalization can be applied to these results to show that what appears to be rapidly varying transient behavior, under suitable transformation, is actually a stationary front. Here we seek a fixed point of the composition of the coarse timestepper with dynamic rescaling of the results of coarse timestepping [CBGK03]. The sequence involves lifting from macroscopic observables (deformation field) to consistent molecular configurations, short molecular dynamics evolution, restriction to macroscopic observables, and then *rescaling* of space and the macroscopic observable fields; the latter is performed based on an established, time-stepper based methodology using template functions [RKML03]. Upon convergence, the self-similar shape as well as the exponents of the macroscopically apparent explosion can be estimated for both types of self-similar solutions [Barenblatt96].

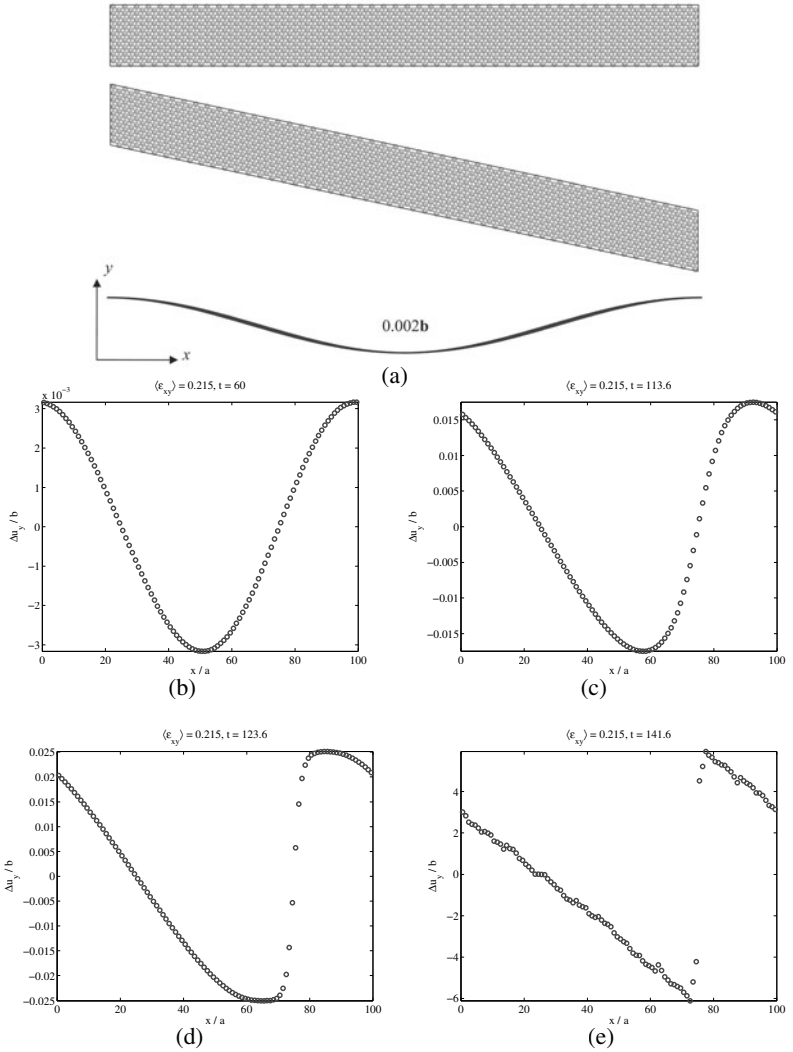


Fig. 4. Visualizing shear strain localization in a crystal under uniform shear deformation. The simulation cell with periodic boundary conditions is shown first undeformed, and then at deformation just before the instability (a), along with an initial sinusoidal shear wave perturbation injected into the system. At the onset of strain localization the perturbation wave profile undergoes four stages of temporal evolution (at a fixed overall strain on the system) that can be classified as linear growth (b), non-linear growth (c), shear-shock formation (d) and formation of an atomic defect, the embryonic dislocation loop / deformation twin (e). After the nucleation, the system strain, previously uniformly distributed, is essentially entirely localized at the glide plane(s) of the dislocation loop / deformation twin.

We have argued here that timestepper-based methods provide a systematic bridge between molecular dynamics timesteppers and traditional continuum numerical analysis. Also that, in the presence of coarse determinism at the level of well-chosen observables, mathematical techniques can be translated into protocols for the judicious design of microscopic computational experiments toward system-level analysis goals. Lifting and restriction protocols between macroscopic observables and microscopic initial conditions provide the dictionary that enables the performance of tasks like coarse fixed point and bifurcation computation, coarse integration, optimization and control, as well as coarse dynamic renormalization. This closure-on-demand approach circumvents the derivation of macroscopic closures; by exploiting the existence of practical determinism, it allows us to solve system-level problems with the least possible extent of microscopic simulation. "On the fly" transfer of information across scales, which relies on variance reduction, filtering and estimation, enables the performance of tasks like reverse integration or the location of transition states, that would be essentially impossible through direct temporal simulation.

Acknowledgment: Support by Honda R&D Co., Ltd. and the OSU Transportation Research Endowment Program (JL), NSF, ONR and Lawrence Livermore National Laboratory (SY) and NSF and AFOSR (YGK) is gratefully acknowledged.

References

[SK93] Shroff, G.M., Keller, H.B.: Stabilization of unstable procedures - the recursive projection method. SIAM Journal on Numerical Analysis, **30**, 1099–1120 (1993)

[TB00] Tuckerman, L.S., Barkley, D: Bifurcation analysis for timesteppers. In: Doedel, E., Tuckerman, L.S. (ed) Numerical Methods for Bifurcation Problems and Large-Scale Dynamical Systems. IMA Volumes in Mathematics and its Applications, **119**, 453-466 (2000)

[TQK00] Theodoropoulos, C., Qian, Y.-H., Kevrekidis, I.G.: "Coarse" stability and bifurcation analysis using timesteppers: a reaction diffusion example. Proc. Natl. Acad. Sci. USA, **97**, 9840–9843 (2000)

[GKT02] Gear, C.W., Kevrekidis, I.G., Theodoropoulos, C.: 'Coarse' integration/bifurcation analysis via microscopic simulators: micro-Galerkin methods. Comput. Chem. Eng., **26**, 941–963 (2002)

[RTK02] Runborg, O., Theodoropoulos, C., Kevrekidis, I.G.: Effective bifurcation analysis: a time-stepper-based approach. Nonlinearity, **15**, 491–511 (2002)

[MMPK02] Makeev, A.G., Maroudas, D., Panagiotopoulos, A.Z., Kevrekidis, I.G.: Coarse bifurcation analysis of kinetic Monte Carlo simulations: A lattice-gas model with lateral interactions. J. Chem. Phys., **117**, 8229–8240 (2002)

[KGHKRT03] Kevrekidis, I.G., Gear, C.W., Hyman, J.M., Kevrekidis, P.G., Runborg, O., Theodoropoulos, C.: Equation-free multiscale computation: enabling microscopic simulators to perform system-level tasks. Comm. Math. Sci., in press (2003). arXiv.org/physics/0209043

[LVZYS02] Li, J., Van Vliet, K.J., Zhu, T., Yip, S., Suresh, S.: Atomistic mechanisms governing elastic limit and incipient plasticity in crystals. Nature, **418**, 307–310 (2002)

[VLZYS03] Van Vliet, K.J., Li, J., Zhu, T., Yip, S., Suresh, S.: Quantifying the early stages of plasticity through nanoscale experiments and simulations. Phys. Rev. B, **67**, 104105 (2003)

[ZLVOYS04] Zhu, T., Li, J., Van Vliet, K.J., Ogata, S., Yip, S., Suresh, S.: Predictive modeling of nanoindentation-induced homogeneous dislocation nucleation in copper. J. Mech. Phys. Solids, (in press)

[TOP96] Tadmor, E.B., Ortiz, M., Phillips, R.: Quasicontinuum analysis of defects in solids. Philos. Mag. A, **73**, 1529–1563 (1996)

[MT02] Miller, R.E., Tadmor, E.B.: The quasicontinuum method: Overview, applications and current directions. J. Comput.-Aided Mater. Des., **9**, 203–239 (2002)

[HK03] Hummer, G., Kevrekidis, I.G.: Coarse molecular dynamics of a peptide fragment: free energy, kinetics, and long-time dynamics computations. J. Chem. Phys., **118**, 10762-10773 (2003)

[GK04] Gear, C.W., Kevrekidis, I.G.: Computing in the past with forward integration. Phys. Lett. A, in press. nlin.CD/0302055

[Li00] Li, J.: Modeling microstructural effects on deformation resistance and thermal conductivity. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge (2000)

[CBCLY01] Cai, W., Bulatov, V.V., Chang, J.-P., Li, J., Yip, S.: Anisotropic elastic interactions of a periodic dislocation array. Phys. Rev. Lett., **86**, 5727–5730 (2001)

[CBCLY03] Cai, W., Bulatov, V.V., Chang, J.-P., Li, J., Yip, S.: Periodic image effects in dislocation modelling. Philos. Mag. A, **83**, 539-567 (2003)

[HSK04] Haataja, M., Srolovitz, D., Kevrekidis, I.G.: Apparent hysteresis in a driven system with self-organized drag. submitted to Phys. Rev. Lett., arXiv.org/cond-mat/0310473.

[ABK04] Aronson, D.G., Betelu, S.I., Kevrekidis, I.G.: Going with the Flow: a Lagrangian approach to self-similar dynamics and its consequences. Submitted to Proc. Natl. Acad. Sci. USA, arXiv.org/nlin/0111055.

[SKK03] Siettos, C., Kevrekidis, I.G., Kevrekidis, P.G.: Focusing revisited: a renormalization/bifurcation approach. Nonlinearity, **16**, 497–506 (2003)

[RM00] Rowley, C.W., Marsden, J.E.: Reconstruction equations and the Karhunen-Loeve expansion. Physica D, **142**, 1–19 (2000)

[RKML03] Rowley, C.W., Kevrekidis, I.G., Marsden, J.E., Lust, K.: Reduction and reconstruction for self-similar dynamical systems. Nonlinearity, **16**, 1257-1275 (2003)

[CBGK03] Chen, L., Debenedetti, P.G., Gear, C.W., Kevrekidis, I.G.: From molecular dynamics to coarse self-similar solutions: a simple example using equation-free computation. Submitted to J. Non-Newtonian Fluid Mech., Sept. 2003.

[Chang03] Chang, J.-P.: Atomistics of defect nucleation and mobility: dislocations and twinning. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge (2003)

[Barenblatt96] Barenblatt, G.I.: Scaling, self-similarity, and intermediate asymptotics. Cambridge University Press, New York (1996)

Exploration of coarse free energy surfaces templated on continuum numerical methods

Daniele Passerone[1] and Ioannis G. Kevrekidis[2]

[1] Computational Laboratory (CoLab), Swiss Federal Institute of Technology (ETH), Hirschengraben 84, 8092 Zurich, Switzerland daniele.passerone@ethz.ch
[2] Chemical Engineering, PACM and Mathematics, Princeton University, Princeton, NJ 08544, USA yannis@princeton.edu

Summary. The characterization of a process in physics or chemistry through a set of collective variables (distances between atoms, orientational order parameters, dihedral angles) can be of great help in understanding the reaction mechanism. In particolar, a key role is played by the free energy surface as a function of these collective variables ("coarse" free energy), and the development of efficient methods of exploration of this surface in order to find minima and saddle points is an important issue. To this end, for example, the ideas of metadynamics were recently proposed [LP02, ILP03]. In this paper we propose a method for exploring the coarse free energy surface limiting the necessary calculations as much as possible, without explicitely resorting to a coarse non-Markovian metadynamics, and using the ideas of continuum numerical methods. We choose as a test case the free energy surface of the 38 atoms Lennard-Jones cluster, which presents an interesting double-funnel structure.

1 Introduction

In a recent paper, Hummer and Kevrekidis [HK03] described a "coarse molecular dynamics" scheme based on a small set of *observation variables*. The dynamics of these coarse observation variables are explored through short bursts of appropriately initialized multiple replica simulations over short times. The result of these simulations allows one to estimate the so-called *coarse timestepper* (alternatively, estimate the time derivative of closed dynamic equations for the expected evolution of the coarse variables). A *meaningful* coarse timestepper provides a bridge between fine scale, atomistic simulation (here MD simulation) and traditional continuum numerical analysis. Algorithms like coarse projective integration forward and backward in time, coarse fixed point solvers and eigensolvers based on matrix-free linear algebra, as well as gaptooth and patch-dynamics ideas for spatially distributed systems can be wrapped around the MD simulator and used to efficiently explore macroscopic, expected dynamics through "on demand" MD computation [TQK00, GKT02, MMK02, GK03a, SGK03, LKG03, GLK03, GHK02].

In this approach the fine-scale, MD code can be thought of as a (computational) experiment. Since this experiment can be initialized and run at will, it can be used

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

to *estimate* the quantities that closed macroscopic equations would give by simple evaluation. Traditional numerical algorithms can be thought as protocols for where to evaluate a model formula, and how to process the results of these evaluations towards an ultimate user-specified goal. In this sense, our coarse grained approach can be thought of as a protocol for designing (computational) experiments with the MD code, and processing their results towards the same goal. The goal in this case is the efficient computational exploration of free-energy surfaces.

There are two important assumptions underlying this computer-assisted analysis approach; the first is that a low-dimensional attracting "slow manifold" does indeed underlie the evolution of expected system dynamics. The second is that the variables chosen are meaningful in parametrizing this manifold (i.e. the manifold does not "fold over" them). The first assumption (low-dimensionality of the expected dynamics, at some level of observation) is related to the very premise that modeling a system is useful: if this is not the case, one cannot be practically predictive about the system in question. The second assumption -*i.e.*, the identification of good reaction coordinates- is a much more difficult one. There exist problems (like chemical kinetics, or laminar fluid flow) where long experience asserts the observation variables that one can write macroscopic equations for: concentrations for kinetics, density and momentum fields for fluid flow, in general a few low order moments of microscopically evolving distributions. In the problems we are trying to explore, it is much more difficult to identify good such variables from scratch in a new problem. This can happen, for example, in condensed systems or clusters, as well as in problems with solute-solvent interactions. Several ingenious approaches, from Voter's accelerated dynamics [MV02] to the dimer method of Henkelman and Jonsson [HJ99] and from the nudged elastic band [HJ00] to its modification by E *et al.* [E02] are being developed to tackle these difficult issues.

When the two basic assumptions are not met, *i.e.* when (to use the language of [HK03]) microscopically evolving distributions in phase space do not become quickly effectively low-dimensional, and parametrizable by our chosen observation variables, our computer-assisted procedure will fail. It is possible to handle such difficult problems *only partially*; one might restrict the goals of the research, giving up the true dynamics of the system and focusing instead on the efficient computational exploration of certain features of the coarse free energy surface, such as finding stationary points and transition states.

Within this more modest context, one may try to fit a system into the previous scheme by introducing some collective variables and "forcing" an adiabatic separation from the atomistic variables. Under certain conditions, these collective variables will explore the same effective free energy surface as in the original problem. This can be achieved following a similar scheme to that proposed in [ILP03] by Iannuzzi *et al.*. In their paper, the ideas of the metadynamics proposed in [LP02] (and, in turn, inspired by [TQK00, GKT02]) are applied to the Car-Parrinello molecular dynamics (MD) scheme. The resulting dynamics are not a "true" dynamics, but explore *the same* coarse free energy surface in an efficient way, with a suitable choice for the parameters (see below). In that paper it was shown how a reaction process, over-

coming the accessible timescales of traditional MD, can be greatly accelerated, thus efficiently exploring the reaction pathways on the coarse free energy surface.

The possibility we investigate here is to apply a similar approach, hopefully without the need of a coarse non-Markovian metadynamics repeatedly revisiting regions during exploration of the entire coarse phase space. We focus instead on limiting the necessary calculations as much as possible, exploiting coarse projective integration and coarse fixed point algorithms, and, more generally, continuum numerical methods.

The test problem we propose is the investigation of the free energy surface of the 38 atom Lennard-Jones cluster. This model problem has been chosen by several authors for testing optimization algorithms or novel MD and Monte Carlo (MC) methods because of its highly non-trivial potential energy surface. It is well established that the system is characterized by a double-funnel free energy landscape, and is thus representative of an important class of complex systems in physics and biochemistry. One funnel ends in the global minimum, a face-centered-cubic (fcc) truncated octahedron, whereas at the bottom of the other funnel is the second lowest energy minimum which is an incomplete Mackay icosahedron (figure 1).

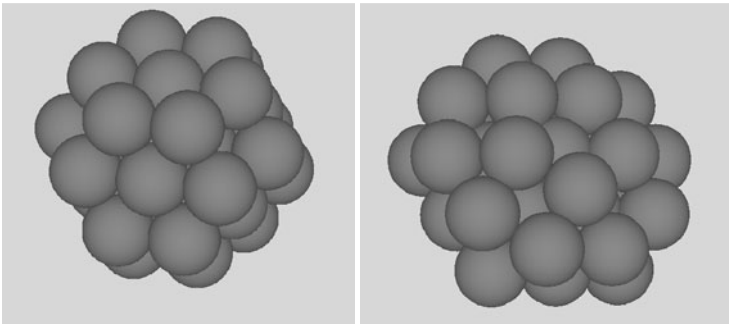


Fig. 1. The two lowest energy minima for the 38-atom Lennard-Jones cluster: a face centered cubic minimum (left) and a Mackay icosahedron (right). (See Plate 9 on Page 273)

Doye, Miller and Wales [DMW99] characterize the energy landscape first by building a disconnectivity graph for the lowest energy minima; then they compute the free energy as a function of a bond-order parameter. A combined application of these two tools allows them to explain why the icosahedral funnel is more accessible when relaxing down the energy landscape: in particular, there is a similarity between the liquidlike state and the icosahedral structure, and a high free energy barrier between the fcc and the icosahedral funnel, trapping the system in one of the funnels at low temperature.

In our work, instead of using a single order parameter for parametrizing the free energy, we decided to use two distinct order parameters simultaneously (the fourfold

and the sixfold orientational order parameters) as collective variables, with the aim of better distinguishing different states of the system on this two-dimensional free energy surface. In this short report, we will briefly describe the techniques used in our method; quantitative results will appear in a following paper [PK04].

2 Physical Problem/ Governing Equations

Our multiscale method involves a microscopic, direct simulation level and a coarse, observation level. The former obeys equations of motion derived from an extended Lagrangian, described below. The latter will be explored through projective integration techniques, and will be described in the following section.

2.1 Extended Lagrangian

We consider a classical system of N particles ($3 \times N$ degrees of freedom) interacting through a potential $V(\mathbf{r}_1, \ldots, \mathbf{r}_N)$. Suppose that this system has at least two stable basins (the "reactants" basin and the "products" basin). Moreover, a set of collective variables $S_\alpha(\mathbf{r})$ which are able to discriminate between reactants and products is selected. These collective variables can be distances between atoms, dihedral angles, coordination numbers, mode coupling variables, or any other function of the atomic coordinates $\mathbf{r}_i$. To be "meaningful", as we discussed above, they should include the relevant modes responsible for the transition from one basin to another; such modes are in principle not easily discernible in traditional MD simulations.

One of the main ideas proposed in [ILP03] is to treat the S_α as dynamical variables (called s_α) and to couple the latter with their instantaneous realization $S_\alpha(\mathbf{r})$ (much in the spirit of the Parrinello-Rahman [PR81] MD scheme). The Lagrangian is then:

$$L = \frac{1}{2}\sum_{i=1}^{N} m_i \left|\dot{\mathbf{r}}_i\right|^2 + \frac{1}{2}\sum_{\alpha} M_\alpha \dot{s}_\alpha^2 - V(\mathbf{r}) - \frac{1}{2}\sum_{\alpha} K_\alpha \left(S_\alpha(\mathbf{r}) - s_\alpha\right)^2 . \qquad (1)$$

In this Lagrangian m_i are the particle masses, K_α is an elastic constant used to keep the instantaneous value of $S_\alpha(\mathbf{r})$ close to the dynamical variable s_α, and M_α is the (fictitious) mass of the variable s_α. In our version of the method we will not add a history-dependent term as in [LP02] and [ILP03]. We will rather use constrained or restrained dynamics for prescribing particular values of the dynamical variables. After that, we will create independent copies of the system compatible with this prescription, and we will let them evolve freely according to (1) over short times. We will thus be able to estimate the "coarse timestepper" of *these* dynamics, or, alternatively, the time derivatives of the coarse variable evolution. These derivatives can serve as input for projective integration / fixed point schemes aimed at finding stationary points on the coarse free energy surface.

It must be noted that the s_α will evolve according to the forces $\Phi_\alpha = K_\alpha(S_\alpha(\mathbf{r}) - s_\alpha)$. In the limit of adiabatic separation between these variables and the atomic degrees of freedom (*i.e.*, in the limit of very large M_α), s_α is approximately fixed while Φ_α is dynamically averaged over the atomic degrees of freedom. As pointed out in [ILP03], s_α will in this case evolve with the force $< \Phi_\alpha >$, that is, the derivative of the free energy with respect to s_α as in standard umbrella sampling or constrained MD.

2.2 Lennard-Jones case

In this work we will focus on a Lennard-Jones cluster, whose atoms interact through the distance-dependent potential

$$V_{LJ}(r) = 4\epsilon \left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right). \tag{2}$$

In this potential (mimicking the hard core repulsion and the induced dipole-induced dipole attraction), which is suitable for noble gas systems like Argon, ϵ is the depth of the well and σ is the hard sphere diameter. The lowest lying energy state of a cluster of 38 atoms has a fcc-like structure, whereas the second lowest minimum has the structure of a Mackay icosahedron.

2.3 Order parameters

Bond-order parameters, which were initially introduced by Steinhardt *et al.* [SNR85], have been used to characterize the free energy barrier for the nucleation of a crystal from a melt because they can differentiate between the fcc crystal and the liquid.

The definition of the order parameter Q_l is:

$$Q_l = \left(\frac{4\pi}{2l+1} \sum_{m=-l}^{l} \left|\bar{Q}_{lm}\right|^2 \right)^{1/2}, \tag{3}$$

where

$$\bar{Q}_{lm} = \frac{1}{\sum_{i'j'} f(r_{i'j'}, r_0)} \sum_{ij} f(r_{ij}, r_0)\, Y_{lm}(\theta_{ij}, \phi_{ij}), \tag{4}$$

where the sum is over all the atoms, $f(r, r_0)$ is a smooth Fermi-like function which is (almost) zero for $r > r_0$, used to select the nearest neighbors (we used $r_0 = 1.391\sigma$); $Y_{lm}(\theta, \phi)$ is a spherical harmonic and θ_{ij} and ϕ_{ij} are the polar and azimuthal angles of an interatomic vector with respect to an arbitrary coordinate frame. We must stress that Q_l is independent of the coordinate frame chosen.

Doye *et al.* [DMW99] show that the Q_4 values of the two lowest energy minima are well-separated. In particular the minimum energy path for the transition from fcc to icosahedral goes from $Q_4 = 0.19$ for the fcc to Q_4 about zero for the icosahedral minimum (see fig. 4 in [DMW99]). Consequently, they used this order parameter for calculating the (very high) free energy barrier between these two classes of minima.

Nevertheless, Q_4 is not able to distinguish between liquidlike states and icosahedral states, since the local fourfold order of these configurations is very similar. On the other hand, the probability distribution of the sixfold order parameter Q_6 has two *separate* maxima (at temperatures around and below the melting temperature for this cluster) corresponding to liquidlike and icosahedral states. This bimodal distribution corresponds to two distinct minima in the free energy, separated by a small barrier. There are therefore three stable basins, a liquidlike state, an icosahedral state, and a narrow fcc state, difficult to reach upon cooling.

As a test case for our method we choose to use both Q_4 and Q_6 as collective variables s_α; we will attempt (in [PK04]) to locate the different features of this interesting system through a systematic, numerically motivated exploration of the relevant zones of a two-dimensional coarse free energy surface.

2.4 Technical details

In the following we will briefly list a set of technical issues that had to be resolved for this particular system, and the solutions we propose.

Prescribing the values of Q_4 and Q_6

A simple way was to restrain the value of Q_l to a certain value $\bar{Q}_l$ using a harmonic potential similar to that introduced in (1). Since Q_l is very sensitive to atomic vibration, it is very difficult to restrain it in this way: whatever the value of the spring constant, the actual value of the parameter oscillates at about ten percent around its expected value. To avoid this, we implemented a SHAKE-like constrained MD (see [K99] and references therein), thus obtaining more precise values of the desired parameter.

Avoiding rotations

Since non-physical forces are applied to the system when evolving it according to the Lagrangian (1), or when restraining the order parameter to a given value, spurious rotations can be induced in the clusters. This can result into an undesirable distribution of the kinetic energy during the MD simulation. An efficient way to avoid this behavior is to impose the nonholonomic constraint of null angular momentum on the system. To satisfy this constraint along with the other constraints simultaneously imposed on the system within the desired precision, a Lagrange-multiplier scheme first proposed by Kutteh [K99] was implemented. Using this method we could constrain simultaneously:

- the kinetic energy (for a simple temperature tuning - isokinetic ensemble)
- the angular momentum to avoid rigid rotations
- the linear momentum to avoid rigid translations
- Q_6 and Q_4.

This approach to removing spurious rotations appears promising and will be discussed in detail in a forthcoming paper [P03].

Choosing the parameters in the Lagrangian

The criterion here was the transformation of the Q_l into "slow variables" in order to fulfill the assumption of adiabatic separation. The choice of the K_α and M_α in eq. (1) has to satisfy two criteria: the first is that typical time scales of the coarse variables have to be well separated from the ones of the atomic motion. This criterion is met if $\sqrt{M_\alpha/K_\alpha}$ is much larger than the typical microscopic period of oscillation of the atomic degrees of freedom. The second criterion is that the reduced probability distribution sampled by the coarse variables s_α corresponds to the correct free energy. It can be shown rigorously ([L03]) that this is the case for sufficiently large values of K_α.

For our case, a reasonable choice of the parameters is (for all α) $K_\alpha = 3 \times 10^4$ and $M_\alpha = 2 \times 10^5$. We stress that the limit of large M_α is very important, since it corresponds to a case where the s_α remains nearly constant while the fast degrees of freedom oscillate, and the average force on the s_α converges to the derivative of the effective free energy with respect to s_α. Moreover, we expect the dynamics of the s_α to follow a Newtonian behavior, and this must be taken into account when extracting information about the free energy from the unconstrained replica dynamics.

3 Multiscale/Multiresolution Method

The multiscale method we propose to accelerate the exploration of the effective free energy surface is the *coarse projective integration*. In particular, we will explore whether *reverse projective integration* may accelerate the location of saddle-type, transition states. As was illustrated in [HK03] and later proved in the deterministic case in [GK03b], short forward MD simulation results can be used to estimate the time derivative on the slow manifold, and this in turn can be used to take *backward* projective steps on the free energy surface. This repeated "forward integration, estimation, backward projection" approach will converge on unstable, source-type states on the free energy surface; given appropriate conditions on the stable and unstable eigenvalues of transition states, it is actually possible that such an integration will *converge* to certain types of saddle-type states backward in coarse time. The approach can be initialized close to a minimum on the free energy surface, and - in general- algorithms for the computation of stable manifolds in dynamical systems can be exploited in the exploration of free energy surfaces (see the relevant discussion and references in [HK03]). Here we start by describing the procedure for estimating the inputs to the coarse projective integration algorithm.

4 Results

As an initial test for our method (particularly for testing the constraints) we prepared the system using constrained MD on a grid of points spanning a set of values of Q_6 and Q_4. This grid is shown in figure 2. The temperature chosen was $T = 0.122$ in

Lennard Jones units, or $T = 15K$ for Argon. Every point on the grid corresponds to a constrained molecular dynamics simulation given a pair of values for (Q_4, Q_6). Rotations were avoided through our nonholonomic constraint algorithm.

The next step was to choose selected points on the grid, pick N uncorrelated configurations from the constrained trajectory (with N from 20 to 50), and use these configurations as starting points for unconstrained dynamics evolving according to the Lagrangian (1). In Fig. 3 we show the behavior of one of these replicas starting from $(Q_4^0, Q_6^0) = (0.12, 0.35)$. We observe the slow evolution of the coarse variables, and the oscillation of their microscopic realizations. With our choice of parameters, we observed that this oscillation thermalizes exponentially quickly towards the imposed temperature for the atomic degrees of freedom, whereas the coarse variables (meant as dynamical variables - the s_α) will retain a different temperature externally tunable at will. We can now use the replicas to search for minima on the free energy surface. Starting from $(Q_4^0, Q_6^0) = (0.15, 0.35)$ we plot (see Fig. 4) the average of the quantity $Q_4(Q_4^0, Q_6^0, t)$ and $Q_6(Q_4^0, Q_6^0, t)$. We do the same starting from a neighboring point on the grid at $(Q_4^0, Q_6^0) = (0.1, 0.35)$. We observe that $< Q_6 > (t)$ is close to a stationary value, whereas the time derivative of $< Q_4 > (t)$ is negative. We plan to use this information in a way analogous to [HK03] for accelerating the location of (stable and unstable) stationary states on the free energy surface.

The information from the time evolution of the replicas can be used for estimating quantities leading to quantification of the free energy surface, and to provide inputs to projective integration / fixed point algorithms. In [HK03] a first order effective Langevin equation for the behavior of the average of the collective coordinate was postulated; we intend to also explore in [PK04] whether (even given the adiabatic separation of the collective variables here) a second order effective description may be necessary. This will affect the reinitialization of replica simulations after a coarse integration or coarse Newton-Raphson step. The relatively short, unconstrained run of the replica simulation after the constrained preparation step is important in the equation-free context. We expect the unconstrained dynamics to quickly slave the expected behavior of fast degrees of freedom to the slow ones (in other words, allow the system to approach the slow manifold) thus allowing us to estimate the time derivative of the unavailable closed dynamic equation for the slow evolution.

5 Summary and Conclusions

In this short report we have presented some preliminary considerations for modifying coarse molecular dynamics, holding some promise for systems where meaningful slow variables for the description of a low-dimensional free energy surface are not directly available. The report sets the stage for linking this approach with algorithms such as reverse projective integration in the illustrative example of the 38-atom Lennard-Jones cluster. Issues like nonholonomic constraints for avoiding spurious cluster rotations arise in this case, and we showed that they can be addressed efficiently using a constrained MD approach.

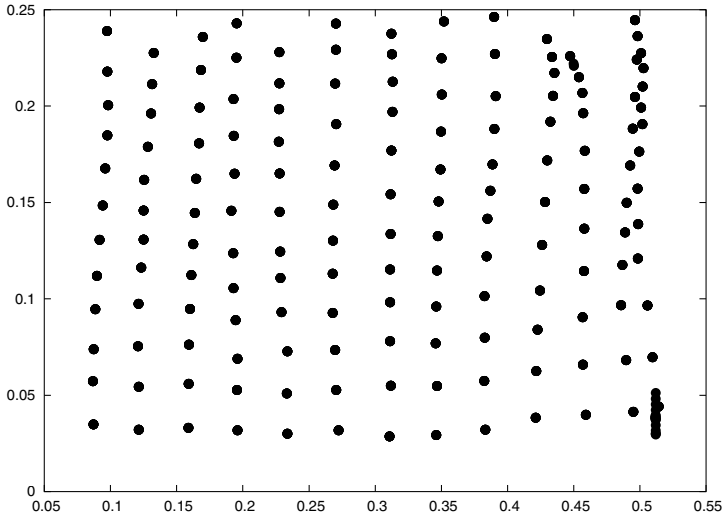


Fig. 2. The grid of points on the coarse free energy surface obtained using constrained MD for the 38-atom Lennard-Jones cluster.

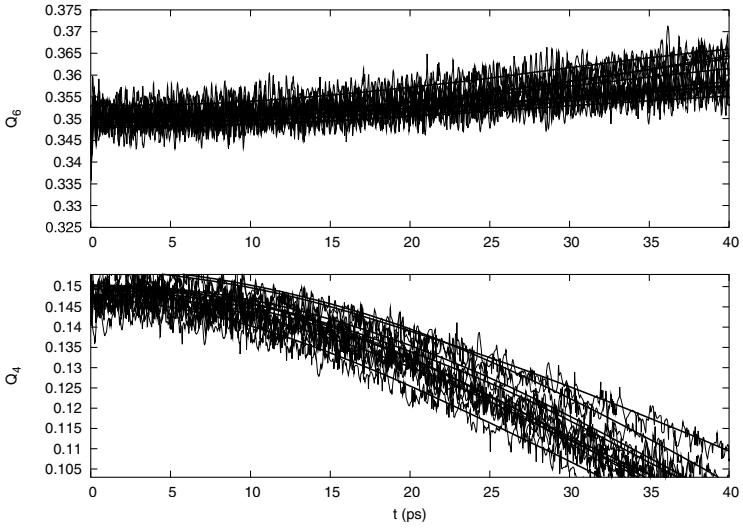


Fig. 3. Q_6 and Q_4 for the unconstrained replica evolution of the system starting from a particular coarse initial condition. The coordinates are forced to be "slow" using large K_α and M_α parameters in eq. (1). The smooth lines correspond to the evolution of the dynamical variables, whereas the oscillating lines represent the instantaneous values of the corresponding order parameters.

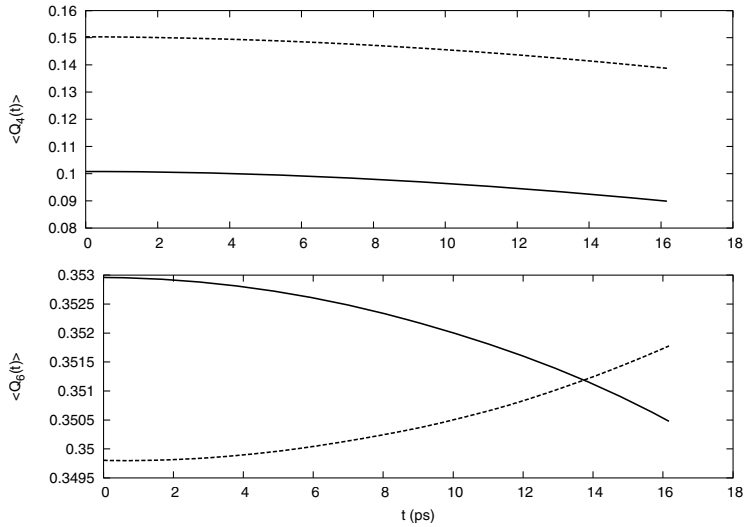


Fig. 4. Average on all replicas ($< Q_4(Q_4^0, Q_6^0, t) >$ and $< Q_6(Q_4^0, Q_6^0, t) >$ for two neighboring starting points on the grid of Fig. 2: (solid) $(Q_4^0, Q_6^0) = (0.1, 0.353)$ and (dashed) $(Q_4^0, Q_6^0) = (0.15, 0.35)$.

D.P. acknowledges illuminating discussions with Alessandro Laio; the work of I.G.K. was partially supported by AFOSR (Dynamics and Control) and an NSF/ITR grant. Both authors acknowledge the hospitality of the 2003 Multiscale Summer School in Lugano.

References

[TQK00] Theodoropoulos, K., Qian, Y.-H., Kevrekidis, I. G.: "Coarse" stability and bifurcation analysis using timesteppers: a reaction diffusion example, Proc. Natl. Acad. Sci. **97**(18), pp.9840-9843 (2000).

[GKT02] Gear, C. W., Kevrekidis, I. G., Theodoropoulos, C.: "Coarse" Integration/Bifurcation Analysis via Microscopic Simulators: micro-Galerkin methods, Comp. Chem. Eng. **26** pp.941-963 (2002)

[MMK02] Makeev, A. G.,Maroudas, D., Panagiotopoulos, A. Z., Kevrekidis, I. G.: Coarse bifurcation analysis of kinetic Monte Carlo simulations: a lattice gas model with lateral interactions J. Chem. Phys. **117**(18) pp.8229-8240 (2002)

[SGK03] Siettos, C., Graham, M. D., Kevrekidis, I. G.: Coarse Brownian dynamics for nematic liquid crystals: Bifurcation, projective integration and control via stochastic simulation. J. Chem. Phys. **118**(22) pp.10149-10157 (2003)

[LKG03] Li, J., Kevrekidis, P. G., Gear, C. W., Kevrekidis, I. G.: Deciding the Nature of the Coarse Equation through Microscopic Simulation: an Augmented Lagrangian Approach, SIAM MMS **1**(3) pp.391-407 (2003) Can be obtained as physics/0212034 at arXiv.org

[GLK03] Gear, C. W., Li, J., Kevrekidis, I. G.: The gaptooth method in particle simulations. *Physics Letters A*, **3**16 pp.190-195 (2003). Can be obtained as physics/0303010 at arXiv.org

[GHK02] Kevrekidis, I. G., Gear, C. W., Hyman, J. M., Kevrekidis, P. G., Runborg, O., Theodoropoulos, K., Kevrekidis, I. G.: Equation-free multiscale computation: enabling microscopic simulators to perform system-level tasks, Comm. Math. Sciences, August 2002, Can be obtained as physics/0209043 at arXiv.org

[MV02] Montalenti, F., Voter, A. F., Exploiting past visits or minimum-barrier knowledge to gain further boost in the temperature-accelerated dynamics method, J. Chem. Phys. **1**16, 4819 (2002) and references therein.

[HJ99] Henkelman, G., Jònsson, H., A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives, J. Chem. Phys. **1**11, 7010 (1999).

[HJ00] Henkelman, G., Jònsson, H., Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points, J. Chem. Phys. **1**13, 9978 (2000).

[E02] E, W. N., Ren, W. Q., Vanden-Eijnden, E., String method for the study of rare events, Phys. Rev. B **66**, 052301 (2002).

[HK03] Hummer, G., Kevrekidis, I. G.: Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computations. J. Chem. Phys., **118**, 10762–10773 (2003)

[ILP03] Iannuzzi, M., Laio, A., Parrinello, M.: Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics. Phys. Rev. Lett., **90**, 238302-1–238302-4 (2003)

[LP02] Laio, A., Parrinello, M.: Escaping free-energy minima. Proceedings of the National Academy of Sciences of the United States of Americas, **99**, 12562–12566 (2002)

[DMW99] Doye, J. P. K., Miller, M. A., Wales, D. J., The double-funnel energy landscape of the 38-atom Lennard-Jones cluster. J. Chem. Phys., **110**, 6896–6906 (1999)

[PK04] Passerone, D., Kevrekidis, I. G., to be published.

[L03] Laio, A.: private communication.

[PR81] Parrinello, M., Rahman, A.: Polymorphic transitions in single-crystals - a new molecular-dynamics method. J. Appl. Phys. **52**, 7182 (1981).

[SNR85] Steinhardt, P. J., Nelson, D. R., Ronchetti, M.: Bond-Orientational Order in Liquids and Glasses. Phys. Rev. B **28**, 784–805 (1983)

[K99] Kutteh, R.: New methods for incorporating nonholonomic constraints into molecular dynamics simulations. J. Chem. Phys., **111**, 1394–1406 (1999)

[P03] Passerone, D., Iannuzzi, M.: Stop that rolling box! (to be published)

[GK03a] Gear, C. W. and Kevrekidis, I. G.: Projective methods for stiff differential equations: problems with gaps in their eigenvalue spectrum, SIAM J. sci. comp. **2**4, 1091 (2003).

[GK03b] Gear, C. W., Kevrekidis, I. G.: Computing in the Past with Forward Integration, submitted to Physics Letters A, February 2003. Can be obtained as nlin.CD/0302055 at arXiv.org.

Damping factors for the gap-tooth scheme

Giovanni Samaey[1], Ioannis G. Kevrekidis[2], and Dirk Roose[1]

[1] Department of Computer Science, K.U. Leuven, Celestijnenlaan 200A, 3000 Leuven, Belgium. giovanni.samaey@cs.kuleuven.ac.be
[2] Department of Chemical Engineering, PACM and Department of Mathematics, Princeton University, Princeton, USA. yannis.kevrekidis@princeton.edu

Summary. An important class of problems exhibits macroscopically smooth behaviour in space and time, while only a microscopic evolution law is known. For such time-dependent multi-scale problems, the gap-tooth scheme has recently been proposed. The scheme approximates the evolution of an unavailable (in closed form) macroscopic equation in a macroscopic domain; it only uses appropriately initialized simulations of the available microscopic model in a number of small boxes. For some model problems, including numerical homogenization, the scheme is essentially equivalent to a finite difference scheme, provided we define appropriate algebraic constraints in each time-step to impose near the boundary of each box. Here, we demonstrate that it is possible to obtain a convergent scheme without constraining the microscopic code, by introducing buffers that "shield" over relatively short time intervals the dynamics inside each box from boundary effects. We explore and quantify the behavior of these schemes systematically through the numerical computation of damping factors of the corresponding coarse time-stepper, for which no closed formula is available.

1 Introduction

For an important class of multi-scale problems, a separation of scales exists between the (microscopic, detailed) level of description of the available model, and the (macroscopic, continuum) level at which one would like to observe the system. Consider, for example, a kinetic Monte Carlo model of bacterial chemotaxis [SGK03]. A stochastic biased random walk model describes the probability of an individual bacterium to run or "tumble", based on the rotation of its flagellae. Technically, it would be possible to run the detailed model for all space and time, and observe the macroscopic variables of interest, but this would be prohibitively expensive. It is known, however, that, under certain conditions, one can write a closed deterministic model for the evolution of the *concentration* of the bacteria as a function of space and time.

The recently proposed *gap-tooth scheme* [KGK02] can then be used instead of performing stochastic time integration in the whole domain. A number of small boxes (*teeth*), separated by large gaps, are introduced; they qualitatively correspond to mesh points for a traditional, continuum solution of the unavailable chemotaxis equation. In one space dimension, the scheme works as follows. We first choose a

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

number of macroscopic grid points and define a small interval around each grid point; initialize the fine scale, microsopic solver within each interval consistently with the macroscopic initial conditions; and provide each interval with appropriate (as we will see, to some extent artificial) boundary conditions. Subsequently, we use the microscopic model in each interval to simulate evolution until time Δt, and obtain macroscopic information (e.g. by computing the average density in each interval) at time Δt. This amounts to a coarse time-Δt map; this procedure is then repeated.

The generalized Godunov scheme of E and Engquist [EE03] also solves an unavailable macroscopic equation by repeated calls to a microscopic code; however, the assumption is made that the unavailable equation can be written in conservation form. In the gap-tooth scheme discussed here, the microscopic computations are performed without assuming such a form for the "right-hand-side" of the unavailable macroscopic equation; we evolve the detailed model in a subset of the domain, and try to recover macroscopic evolution through interpolation in space and extrapolation in time.

We have showed analytically, in the context of numerical homogenization, that the gap-tooth scheme is close to a finite difference scheme for the homogenized equation [SKR03]. However, that analysis employed simulations using an algebraic constraint, ensuring that the initial macroscopic gradient is preserved at the boundary of each box over the time-step Δt. This may require altering an existing microscopic code, since a given microscopic code typically only allows a set of predefined boundary conditions. It is highly non-trivial to impose macroscopically inspired boundary conditions on such microscopic codes, see e.g. [LLY98]. This can be circumvented by introducing buffer regions at the boundary of each small box, which shield the dynamics within the computational domain of interest from boundary effects over short time intervals. If an appropriate buffer size can be determined, one can subsequently use the microscopic code with its *built-in* boundary conditions.

Here, we show we can study the gap-tooth scheme (with buffers) through its numerically obtained damping factors, by estimating its eigenvalues. For a one-dimensional diffusion problem, we show that the eigenvalues of the gap-tooth scheme are approximately the same as those of the finite difference scheme. When we impose Dirichlet boundary conditions at the boundary of the buffers, we show that the scheme converges to the standard gap-tooth scheme for increasing buffer size.

The model problem was chosen because of its simplicity, since in this case we can compute the damping factors analytically. Of course, the gap-tooth scheme will only be efficient when no macroscopic model is available and the only alternative is to perform a numerical simulation on the whole domain using a very fine mesh. An example is the case of a homogenization problem, for which the scheme has similar behaviour [SKR03].

2 Physical Problem/ Governing Equations

Consider a general reaction-convection-diffusion equation with a dependence on a small parameter ϵ,

$$\frac{\partial}{\partial t}u(x,t) = f\left(u(x,t), \frac{\partial}{\partial x}u(x,t), \frac{\partial^2}{\partial x^2}u(x,t), x, \frac{x}{\epsilon}\right), \tag{1}$$

with initial condition $u(x,0) = u_0(x)$ and Dirichlet boundary conditions $u(0,t) = v_o$ and $u(1,t) = v_1$, and assume that f is 1-periodic in $y = \frac{x}{\epsilon}$.

Since we are only interested in the macroscopic (averaged) behavior, let us define an averaging operator for $u(x,t)$ as follows

$$U(x,t) := \mathcal{S}_h(u)(x,t) = \int_{x-\frac{h}{2}}^{x+\frac{h}{2}} u(\xi,t)\mathrm{d}\xi. \tag{2}$$

This operator replaces the unknown function with its local average in a small box of size $h >> \epsilon$ around each point. If h is sufficiently small, this amounts to the removal of the microscopic oscillations of the solution, retaining its macroscopically varying components.

The averaged solution $U(x,t)$ satisfies an (unknown) macroscopic partial differential equation,

$$\frac{\partial}{\partial t}U(x,t) = F\left(U(x,t), \frac{\partial}{\partial x}U(x,t), \frac{\partial^2}{\partial x^2}U(x,t), x; h\right), \tag{3}$$

which depends on the box width h.

The goal of the gap-tooth scheme is to approximate the solution $U(x,t)$, while only making use of the detailed model (1). For analysis purposes, consider as a microscopic model the constant coefficient diffusion equation,

$$\frac{\partial}{\partial t}u(x,t) = a^*\frac{\partial^2}{\partial x^2}u(x,t), \tag{4}$$

Both $U(x,t)$ and $u(x,t)$ then satisfy (4). The microscopic and macroscopic models are the same, which allows us to focus completely on the method and the properties of its damping factors, see section 4.2.

3 Multiscale/Multiresolution Method

3.1 The gap-tooth scheme

Suppose we want to obtain the solution of the *unknown* equation (3) on the interval $[0,1]$, using an equidistant, macroscopic mesh $\mathit{\Pi}(\Delta x) := \{0 = x_0 < x_1 = x_0 + \Delta x < \ldots < x_N = 1\}$. To this end, consider a small interval (*tooth*, box) of length $h << \Delta x$ centered around each mesh point, and let us perform a time integration

using the microscopic model (1) in each box. We provide each box with boundary conditions and initial condition as follows.

Boundary conditions. Since the microscopic model (1) is diffusive, it makes sense to impose a fixed gradient at the boundary of each small box for a time Δt for the macroscopic function $U(x,t)$. The value of this gradient is determined by an approximation of the concentration profile by a polynomial, based on the (given) box averages U_i^n, $i = 1, \ldots, N$.

$$u(x,t_n) \approx p_i^k(x;t_n), \qquad x \in [x_i - \frac{h}{2}, x_i + \frac{h}{2}],$$

where $p_i^k(x;t_n)$ denotes a polynomial of (even) degree k. We require that the approximating polynomial has the same box averages in box i and in $\frac{k}{2}$ boxes to the left and to the right. This gives us

$$\frac{1}{h}\int_{x_{i+j}-\frac{h}{2}}^{x_{i+j}+\frac{h}{2}} p_i^k(\xi;t_n)\mathrm{d}\xi = U_{i+j}^n, \qquad j = -\frac{k}{2}, \ldots, \frac{k}{2}. \tag{5}$$

One can easily check that

$$\mathcal{S}_h(p_i^k)(x,t_n) = \sum_{j=-\frac{k}{2}}^{\frac{k}{2}} U_{i+j}^n L_{i,j}^k(x), \qquad L_{i,j}^k(x) = \prod_{\substack{l=-\frac{k}{2} \\ l\neq j}}^{\frac{k}{2}} \frac{(x - x_{i+l})}{(x_{i+j} - x_{i+l})} \tag{6}$$

where $L_{i,j}^k(x)$ denotes a Lagrange polynomial of degree k. The derivative of $p_i^k(x,t_n)$ at $x_i \pm \frac{h}{2}$,

$$s_i^{\pm} = \left.\frac{\mathrm{d}}{\mathrm{d}x} p_i^k(x;t_n)\right|_{x_i \pm \frac{h}{2}}, \tag{7}$$

is subsequently used as a Neumann boundary condition. In [SKR03], we show how to enforce the macroscopic gradient to be constant in the presence of fast oscillations.

Initial condition. For the time integration, we must impose an initial condition $\tilde{u}^i(x,t_n)$ in each box $[x_i - \frac{h}{2}, x_i + \frac{h}{2}]$, at time t_n. We require $\tilde{u}^i(x,t_n)$ to satisfy the boundary condition and the given box average. We choose a quadratic polynomial, centered around the coarse mesh point,

$$\tilde{u}^i(x,t_n) \equiv a(x - x_i)^2 + b(x - x_i) + c. \tag{8}$$

Using the constraints $\frac{\mathrm{d}}{\mathrm{d}x}\tilde{u}^i(x_i \pm \frac{h}{2}, t_n) = s_i^{\pm}$ and requiring $\frac{1}{h}\int_{x_i-\frac{h}{2}}^{x_i+\frac{h}{2}} \tilde{u}^i(\xi,t_n)\mathrm{d}\xi = U_i^n$, we obtain

$$a = \frac{s_i^+ - s_i^-}{2h}, \qquad b = \frac{s_i^+ + s_i^-}{2}, \qquad c = U_i^n - \frac{h}{24}(s_i^+ - s_i^-). \tag{9}$$

The algorithm. The complete *gap-tooth* algorithm to proceed from t_n to $t_{n+1} = t_n + \Delta t$ is given below (see figure 1):

1. At time t_n, construct the initial condition $\tilde{u}^i(x, t_n)$, $i = 0, \ldots, N$, using the box averages U_j^n ($j = 0, \ldots, N$) as defined in (9).
2. Compute $\tilde{u}^i(x, t)$ by solving the equation (1) until time $t_{n+1} = t + \Delta t$ with Neumann boundary conditions (7).
3. Compute the box average U_i^{n+1} at time t_{n+1}.

It is clear that this amounts to a "coarse to coarse" time-Δt map. We write this map as follows,

$$U^{n+1} = S_k(U^n; t_n + \Delta t), \tag{10}$$

where S represents the numerical time-stepping scheme for the macroscopic (coarse) variables and k denotes the degree of interpolation.

We emphasize that the scheme is also applicable if the microscopic model is not a partial differential equation by replacing step 2 by a so-called coarse time-stepper [GKT02]. In fact, this is the case for which it was originally designed [KGK02]. Numerical experiments using this algorithm are presented in [GLK03, GK02].

3.2 The gap-tooth scheme with buffers

We already mentioned that, in many cases, it is not possible or convenient to constrain the macroscopic gradient. However, the only crucial issue is that the detailed system in each box should evolve *as if it were embedded in a larger domain*. This can be effectively accomplished by introducing a larger box of size $H >> h$ around each macroscopic mesh point, but still only use (for macro-purposes) the evolution over the smaller, "inner" box. This is illustrated in figure 1. Lifting and evolution (using *arbitrary* outer boundary conditions) are performed in the larger box; yet the restriction is done by taking the average of the solution over the inner, small box. The goal of the additional computational domains, the *buffers*, is to buffer the solution inside the small box from outer boundary effects. This can be accomplished over *short enough* time intervals, provided the buffers are *large enough*; analyzing the method is tantamount to making these statements quantitative.

The idea of using a buffer region was also used in the multi-scale finite element method (oversampling) of Hou [HW97] to eliminate the boundary layer effect; also Hadjiconstantinou makes use of overlap regions to couple a particle simulator with a continuum code [Had99]. If the microscopic code allows a choice of different types of "outer" microscopic boundary conditions, selecting the size of the buffer may also depend on this choice.

4 Results

We first show analytically and numerically that the standard gap-tooth scheme converges for equation (4). We then analyze convergence of the scheme with buffers and Dirichlet boundary conditions through its damping factors.

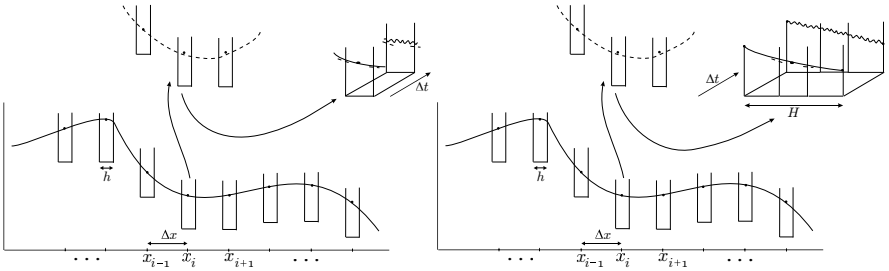


Fig. 1. A schematic representation of the two variants of the gap-tooth scheme (with (left) and without (right) buffer boxes).

4.1 Convergence of the gap-tooth scheme

Theorem 1. *The gap-tooth scheme for equation (4) with analytical integration within the boxes, and boundary conditions defined through interpolating polynomials of (even) order k, is equivalent to a finite difference discretization of order k central differences in space and an explicit Euler time step.*

Proof. When using exact (analytic) integration in each box, we can find an explicit formula for the gap-tooth time-stepper. The initial profile is given by (8), $\tilde{u}^i(x,t_n) = a(x-x_i)^2+b(x-x_i)+c$. Due to the Neumann boundary conditions, time integration can be done analytically, using

$$\tilde{u}^i(x,t_n+\Delta t) = a(x-x_i)^2 + b(x-x_i) + c + 2a\cdot a^*\Delta t.$$

Averaging this profile over the box gives the following time-stepper for the box averages,

$$U_i^n = U_i^n + a^*\frac{s_i^+ - s_i^-}{h}\Delta t.$$

We know that $s_i^\pm = \mathrm{D}_x(p_i^k)(x_i\pm h/2,t_n)$, where $p_i^k(x,t_n)$ is determined by (5). One can easily verify that $\mathrm{D}_x^2\mathcal{S}_h(p_i^k)(x,t_n) = (s_i^+ - s_i^-)/h$, is a k-th order approximation of $\partial_x^2 u$, which concludes the proof. □

As an example, we apply the gap-tooth scheme to the diffusion equation (4) with $a^* = 1$. We choose an initial condition $U(x,t) = 1-|2x-1|$, with Dirichlet boundary conditions, and show the result of a fourth-order gap-tooth simulation with $\Delta x = 0.05$, $\Delta t = 5\cdot 10^{-3}$ and $h = 0.01$. Inside each box, we used a second order finite difference scheme with microscopic spatial mesh size $\delta x = 1\cdot 10^{-3}$ and $\delta t = 5\cdot 10^{-7}$. The results are shown in figure 2.

4.2 Damping factors

Convergence results are typically established by proving consistency and stability. If one can prove that the error in each time step can be made arbitrarily small by refining the spatial and temporal mesh size, and that an error made at time t_n does

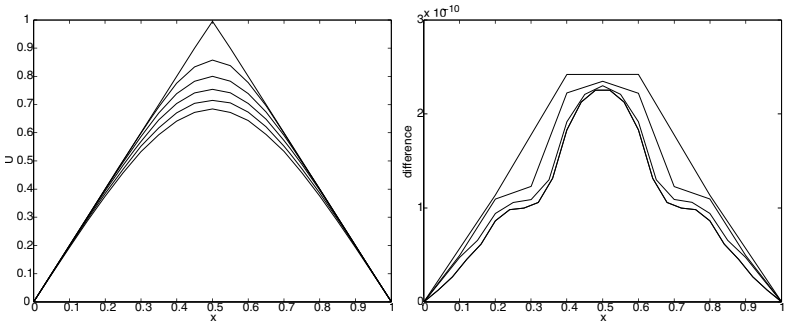


Fig. 2. The gap-tooth scheme of fourth order for eq. (4) at $t = 0, 4 \cdot 10^{-3}, \ldots, 2 \cdot 10^{-2}$.

not get amplified in future time-steps, one has proved convergence. This requires the solution operator to be stable as well.

In the absense of explicit formulas, one can examine the damping factors of the time-stepper. If, for decreasing mesh sizes, all (finitely many) eigenvalues and eigenfunctions of the time-stepper converge to the dominant eigenvalues and eigenfunctions of the time evolution operator, one expects the solution of the scheme to converge to the true solution of the evolution problem.

Consider equation (4) with Dirichlet boundary conditions $u(0,t) = 0$ and $u(1,t)$, and denote its solution at time t by the time evolution operator

$$u(x,t) = s(u_0(x);t), \tag{11}$$

We know that

$$s(\sin(m\pi x);t) = \mathrm{e}^{-(m\pi)^2 t}\sin(m\pi x), \qquad m \in \mathbb{N}.$$

Therefore, if we consider the time evolution operator over a fixed time $\bar{t}$, $s(\cdot,\bar{t})$, then this operator has eigenfunctions $\sin(m\pi x)$, with resp. eigenvalues

$$\lambda_m = \mathrm{e}^{-(m\pi)^2 \bar{t}}. \tag{12}$$

A good (finite difference) scheme approximates well all eigenvalues whose eigenfunctions can be represented on the given mesh. We choose $\bar{t}$ a multiple of Δt for convenience.

Since the operator defined in (11) is linear, the numerical time integration is equivalent to a matrix-vector product. Therefore, we can compute the eigenvalues using matrix-free linear algebra techniques, even for the gap-tooth scheme, for which it might not even be possible to obtain a closed expression for the matrix. We note that this analysis gives us an indication about the quality of the scheme, but it is by no means a proof of convergence.

We illustrate this with the computation of the eigenvalues of the gap-tooth scheme with Neumann box boundary conditions. In this case, we know from theorem 1 that these eigenvalues should correspond to the eigenvalues of a finite difference scheme on the same mesh. We compare the eigenvalues of the second-order

gap-tooth scheme for equation (4) with diffusion coefficient $a^* = 0.45825686$. As method parameters, we choose $\Delta x = 0.05$, $h = 5 \cdot 10^{-3}$, $\Delta t = 2.5 \cdot 10^{-4}$ for a time horizon $\bar{t} = 4 \cdot 10^{-3}$, which corresponds to 16 gap-tooth steps. Inside each box, we use a finite difference scheme of order 2 with $\delta x = 1 \cdot 10^{-4}$ and an implicit Euler time-step of $5 \cdot 10^{-5}$. We compare these eigenvalues to those of the finite difference scheme with $\Delta x = 0.05$ and $\Delta t = 2.5 \cdot 10^{-4}$, and with the dominant eigenvalues of the "exact" solution (a finite difference approximation with $\Delta x = 1 \cdot 10^{-3}$ and $\Delta t = 1 \cdot 10^{-7}$). The result is shown in figure 3. The difference between the finite difference approximation and the gap-tooth scheme in the higher modes, which should be zero according to theorem 1, is due to the numerical solution inside each box and the use of numerical quadrature for the average.

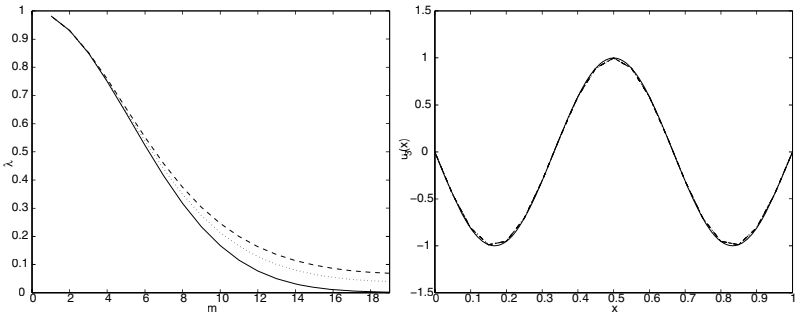


Fig. 3. Comparison between the damping factors (left) and the eigenfunction $u_3(x)$ corresponding to eigenvalue λ_3 (right) of the exact solution (full line), the finite difference approximation (dashed) and the gap-tooth scheme (dotted).

We now examine the effect of introducing a buffer region, as described in section 3.2. We consider again equation (4) with $a^* = 0.45825686$, and we take the gap-tooth scheme with parameters $\Delta x = 0.05$, $h = 5 \cdot 10^{-3}$, $\Delta t = 2.5 \cdot 10^{-4}$ for a time horizon $\bar{t} = 4 \cdot 10^{-3}$, and an internal time-stepper as above. We introduce a buffer region of size H, and we impose Dirichlet boundary conditions at the outer boundary of the buffer region. Lifting is done in identically the same way as for the gap-tooth scheme without buffers; we only use (9) as the initial condition in the larger box $[x_i - \frac{H}{2}, x_i + \frac{H}{2}]$. We compare the eigenvalues again with the eigenvalues of the equivalent finite difference scheme and those of the exact solution, for increasing sizes of the buffer box H. Figure 4 shows that, as H increases, the eigenvalues of the scheme converge to those of the original gap-tooth scheme. We see that, in this case, we would need a buffer of size $H = 4 \cdot 10^{-2}$, i.e. 80% of the original domain, for a good approximation of the damping factors. One could decrease the needed buffer size by decreasing Δt, which results in more re-initializations.

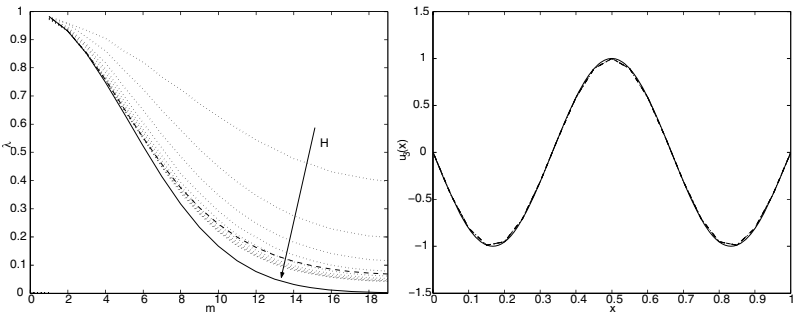


Fig. 4. Comparison between the damping factors (left) and the eigenfunction $u_3(x)$ corresponding to the eigenvalue λ_3 (right) of the exact solution (full line), the finite difference scheme (dashed) and the gap-tooth scheme with buffers (dotted lines) for increasing buffer sizes $H = 2 \cdot 10^{-2}, 3 \cdot 10^{-2} \ldots, 1 \cdot 10^{-1}$.

5 Summary/Conclusions

We described the gap-tooth scheme for the numerical simulation of multi-scale problems. This scheme simulates the macroscopic behaviour over a macroscopic domain when only a microscopic model is explicitly available. In the case of one dimensional diffusion, we showed equivalence of our scheme to standard finite differences of arbitrary (even) order, both theoretically and numerically.

We showed that it is possible, even without analytic formulas, to study the properties of the gap-tooth scheme and generalizations through the damping factors of the resulting coarse time-Δt map. We illustrated this for the original gap-tooth scheme and for an implementation using Dirichlet boundary conditions in a buffer box. We showed that, as long as the buffer region is "large enough" to shield the internal region from the boundary effects over a time Δt, we get a convergent scheme. Therefore, we are able to use microscopic codes in the gap-tooth scheme *without modification*.

The diffusion problem examined here was used as an illustration, because one can compute the damping factors analytically. In a forthcoming paper, we will use these damping factors to study the the trade-off between the effort required to impose a particular type of boundary conditions and the efficiency gain due to smaller buffer sizes and/or longer possible time-steps before reinitialization in the context of homogenization.

Acknowledgements GS is a Research Assistant of the Fund of Scientific Research - Flanders. This work has been partially supported by grant IUAP/V/22 and by the Fund of Scientific Research through Research Project G.0130.03 (GS, DR), and by the AFOSR and the NSF (IGK). The authors thank Olof Runborg for discussions that improved this text and the organizers of the Summer School in Multi-scale Modeling and Simulation in Lugano.

References

[EE03] W. E and B. Engquist. The heterogeneous multi-scale methods. Comm. Math. Sci., 1(1):87–132, 2003.

[GK02] C.W. Gear and I.G. Kevrekidis. Boundary processing for Monte Carlo simulations in the gap-tooth scheme. physics/0211043 at arXiv.org, 2002.

[GKT02] C.W. Gear, I.G. Kevrekidis, and C. Theodoropoulos. "Coarse" integration/bifurcation analysis via microscopic simulators: micro-Galerkin methods. Comp. Chem. Eng., 26(7-8):941–963,2002.

[GLK03] C.W. Gear, J. Li, and I.G. Kevrekidis. The gap-tooth method in particle simulations. Physics Letters A, 316:190–195, 2003.

[Had99] N. G. Hadjiconstantinou. Hybrid atomistic-continuum formulations and the moving contact-line problem. J. Comp. Phys., 154:245–265, 1999.

[HW97] T.Y. Hou and X.H. Wu. A multiscale finite element method for elliptic problems in composite materials and porous media. J. Comp. Phys., 134:169–189, 1997.

[KGK02] I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidis, O. Runborg, and C. Theodoropoulos. Equation-free multiscale computation: enabling microscopic simulators to perform system-level tasks. Comm. Math. Sci. Submitted, physics/0209043 at arxiv.org, 2002.

[LLY98] J. Li, D. Liao, S. Yip. Imposing field boundary conditions in MD simulation of fluids: optimal particle controller and buffer zone feedback, Mat. Res. Soc. Symp. Proc, 538(473-478), 1998.

[SKR03] G. Samaey, I.G. Kevrekidis, and D. Roose. The gap-tooth scheme for homogenization problems. SIAM MMS, 2003. Submitted.

[SGK03] S. Setayeshar, C.W. Gear, H.G. Othmer, and I.G. Kevrekidis. Application of coarse integration to bacterial chemotaxis. SIAM MMS. Submitted, physics/0308040 at arxiv.org, 2003.

Part II

Materials Science

Multiscale Aspects of Polymer Simulations

Kurt Kremer

Max Planck Institute for Polymer Research, 55021 Mainz, Germany,
kremer@mpip-mainz.mpg.de

1 Introduction

Polymers comprise a class of modern synthetic or biogenic materials, which we encounter permanently in our daily life. Thus, a thorough and detailed understanding of their properties beyond certain very general aspects is highly desirable. Since modern research and applications go significantly beyond the traditional view on "simple" polymers such as long chain molecules composed of many identical repeat units, the term polymer science is more and more replaced by soft matter science, indicating not only the huge variety of systems available but also the fact that typical energies or better energy densities are rather small and therefore thermal fluctuations are very important. Still the investigation of various chain-like polymers is an important starting point for the study of more elaborate chemical architectures.

Polymeric materials can be crystalline, amorphous (glasses, rubbers, melts, gels)or even solutions. Especially polymer melts in the glassy state are standard materials for many applications (yoghurt cups, compact discs, housings of technical equipment etc.). They often combine relatively low specific weight and ductibility with processing at moderate temperatures. In the melt state, polymers are viscoelastic liquids where the crossover from elastic to viscous behavior can be adjusted or is determined by the local chemical structure and the chain length. Added to a solvent, polymers can be used as viscosity modifiers and, depending on parameters, be either shear thickening or shear thinning, as used e.g. for drag reduction. Crosslinking chains into a disordered network results in gels or rubber. Applications range from gels in (low fat) food, hydrogels in modern body care (nappies ...) via biological systems (cytoskeleton...) all the way to classical elastomers (e.g. car tires), to name very few. Here, the interplay of connectivity, chain length as well as local structure parameters, such as stiffness determines the properties.

This versatility of physical properties is based on the many different chemical molecular building blocks as well as on various molecular architectures and huge differences in molecular weights of polymers. It is the combination and the rather delicate interplay of local chemical with more global architectural and size properties, which makes macromolecules so versatile and interesting. This means that many

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

different length and time scales are relevant, and that understanding the properties on one scale is not at all sufficient to understand material properties.

The simplest polymers are chain molecules with identical chain segments, repeat units or monomers. To give a very few examples, there are

PE	$(CH_2)_N$	polyethylene
PS	$(CH_2(CH(C_6H_5))_N$	polystyrene
PEO	$((CH_2)_2O)_N$	polyethylene oxide
BPA-PC	$((C_6H_4)C(CH_3)_2(C_6H_4)CO_3)_N$	bisphenol A polycarbonate

from the widely used PE (e. g. plastic bags) to the more complicated, but technically very relevant polycarbonate BPA-PC (compact discs). Many other cases exist, which can be very complex as biological examples demonstrate (DNA, proteins), where several different building blocks are present in one huge molecule. While most polymers are not water soluble, PEO has the exceptional property, that it is both water and oil soluble. Other important water soluble polymers are polyelectrolytes (e. g. sulfonated PS: NaPSS), which in water dissociate into ions and due to this are soluble even though their backbone is hydrophobic. Though not in the focus of the present contribution, the typical simulation approaches for polyelectrolytes however are conceptually very similar to the ones discussed here [KK96, Bas00, Bi95, Mo94].

2 Multi-Scales for Polymer Problems

The relation between atomistic structure and material properties is the basic concern of modern material science. Here macromolecular materials offer many particular challenging hurdles, however, similarly rewarding opportunities. The longstanding aim is far beyond standard properties of bulk materials and focuses more and more on surface interface aspects of the relation between structure and function in nanoscopic molecular assemblies (e. g. self healing nano wires [Pe02]). This all implies a thorough understanding on many length and correspondingly time scales ranging from (sub)atomic to macroscopic. This feature has led to enormous success in describing *generic* aspects, through scaling relations and dimensional arguments [Ge79, Do86] as well as simulations based on simplified models such as bead spring chains or even lattice models [Bi96, Bi95]. However, the astounding variety of material properties equally depends on specific differences in molecular structure.

The importance of different time and length scales can be illustrated by the shear viscosity η of a polymer melt. It contains the whole complexity of the problem. If one changes the process temperature of a BPA-PC melt from 500K to 470K, the viscosity rises by a factor of ten. This is a direct result of the local interactions on the atomistic level as it could -in principle- also have been achieved by an equivalent change of the chemical structure of the monomer. (The glass transition temperature T_G of BPA-PC is around 420K and different polymers display huge differences in T_G.) On the other hand, increasing the chain length by a factor of 2 also shifts the viscosity by a factor of ten, since for melts of long chains one observes $\eta \propto N^{3.4}$.

This power law is a universal property of linear polymers, and holds for all known polymers independent of the chemical structure of the backbone. Thus in the simplest case of linear polymers one can write

$$\eta = AN^{3.4}, \tag{1}$$

which separates the problem into a material specific contribution "A" based on local properties determined on scales of Å or nm size and a generic contribution $N^{3.4}$. Besides this separation feature it is important to mention that both contributions can be the origin of viscosity variations over many decades. A second example concerns the miscibility of polymer blends. In the amorphous state polymer chains are well characterized by a random coil conformation, i. e. $< R^2 > \propto N$. R being the overall chain extension and N being the number of repeat units. Since the material density is homogeneous the volume of each chain is shared by $N^{1/2}$ other chains. Considering the simplest case of a mixture of A-type and B-type polymers of the same length. Then each A chain roughly encounters $0(N)$ AA and $0(N)$ AB contacts along its contour. The same of course holds for the B chains. With typical energies of ϵ_{AA}, ϵ_{AB}, ϵ_{BB}, for AA, AB and BB contacts respectively, the overall interactions energy of a chain of type A is given by

$$U_{AB} \propto N\epsilon_{AB}, U_{AA} \propto N\epsilon_{AA}, U_{BB} \propto N\epsilon_{BB} \tag{2}$$

Since the conformational entropy (cf. Fig. 1) in the AB mixture and the pure A or B phase is about the same, very small differences of the order of k_BT/N in the difference $|\epsilon_{AA} - \epsilon_{AB}|$ are sufficient to drive a phase separation.A striking illustration of this fact is, that even protonated and deuterated polystyrene phase separate, provided the chains are long enough [Ge92]. Again one can manipulate the stability of a blend by changing local interactions (e. g. via the temperature T) or by varying the chain length. This enormous sensitivity on specific local interactions for macroscopic properties marks a special challenge in this field.

At first sight, it might be tempting to perform an all-atom computer simulation of a melt of polymers in order to determine properties like viscosity, morphology etc. However, there are two major complications. The first stems from the choice of interaction potentials, while the second is related to the many scales involved, which is the main focus of the present workshop. An all atom simulation requires the use of an empirical force field. All quantum simulations (Car-Parinello Density Functional simulations, path integral quantum Monte Carlo or combinations thereof) are still confined to very small systems and orders of magnitude slower than force field approaches [Bi96, KK96]. The precondition for such an approach is an empirical energy function for the interaction of all atoms in the system. This determines the force field to solve Newton's equations of motion for the system. Though conceptually straight forward, such an attempt contains a number of complications. First, though usually not considered, are quantum effects. One might think that typical temperatures for macromolecular systems (room temperature and higher) are well above the Debye temperature of the relevant atoms. This is true for the carbon atoms, however, not necessarily for the many hydrogens present. This should be kept in mind

as a general sign of caution. Secondly, for the force field the intra-molecular interactions can be derived from a proper parametrization of quantum calculations on chain fragments. However to parameterize the inter-molecular or non bonded interactions, usually experimental quantities like the heat of vaporization of low-molecular weight liquids are used. There, additional difficulties can arise from the quality and availability of experimental data. Moreover, it is often impossible to optimize all properties to the same degree of accuracy and confidence. This has to be seen also with respect to the example of mixtures discussed before. Thus, one has to be very careful and there is NO single force field for a system which, without further verification, can be used at significantly different temperatures or compositions. In principle, a new parametrization has to be performed for every new set of simulation parameters. Keeping this in mind, force field simulations are very useful and have provided important insight into microscopic properties. Typical examples can be found in the overviews [Bas00, Mo94].

Whether such a fully atomistic simulation, if possible, would be useful at all can be very questionable, since it would provide an enormous amount of data. Almost all the generated information would be irrelevant for a typical problem like the above mentioned viscosity η. In order to make suggestions for material improvements, or to qualitatively and quantitatively understand certain properties, it is crucial to structure and properly interpret the results rather than just collect huge amounts of data. This is often easier and more direct with simplified models.

This leads us to the main topic of this contribution. Polymers can be characterized by a hierarchy of different length and especially time scales, which span a wide range as sketched in. Fig. 1.

On the microscopic level the properties are dominated by the local vibrations of bond angles and lengths. The typical time is of $10^{-13} sec$ as given in the figure resulting in a simulation time step to integrate Newton's equations of motion of about $10^{-15} sec$. This Å-regime is well characterized by the bond angles and bond lengths resulting from the quantum mechanical energy levels. Up to torsional degrees of freedom, chains are in their vibrational ground states, at typical experimental temperatures. The properties on this level are solely determined by the detailed chemical structure of the molecules involved.

On a more coarse grained level, one cannot resolve all the atomistic details of the chains any more. The chain looks like a more or less flexible thread. This is the universal, entropy-dominated coil regime. The many possible conformations of the chains and the many ways to pack chains in a melt or solution determine the morphology. For many questions, the intra-chain entropy plays the most important role. However, once two states of similar intra-chain entropy are available (e.g. chains in a mixture) tiny energy differences in the interaction, originating from the microscopic structure, of order $O(k_B T/N)$ are sufficient to change the morphology completely and to drive the phase separation as mentioned before. On the mesoscopic level many properties can be understood on the basis of simple coarse grained (bead spring) models. On the even coarser level, the semi-macroscopic level the behavior is dominated by the overall relaxation of conformations of the objects. The typical relaxation time, depending on chain length varies between $\propto N^2$ for short and

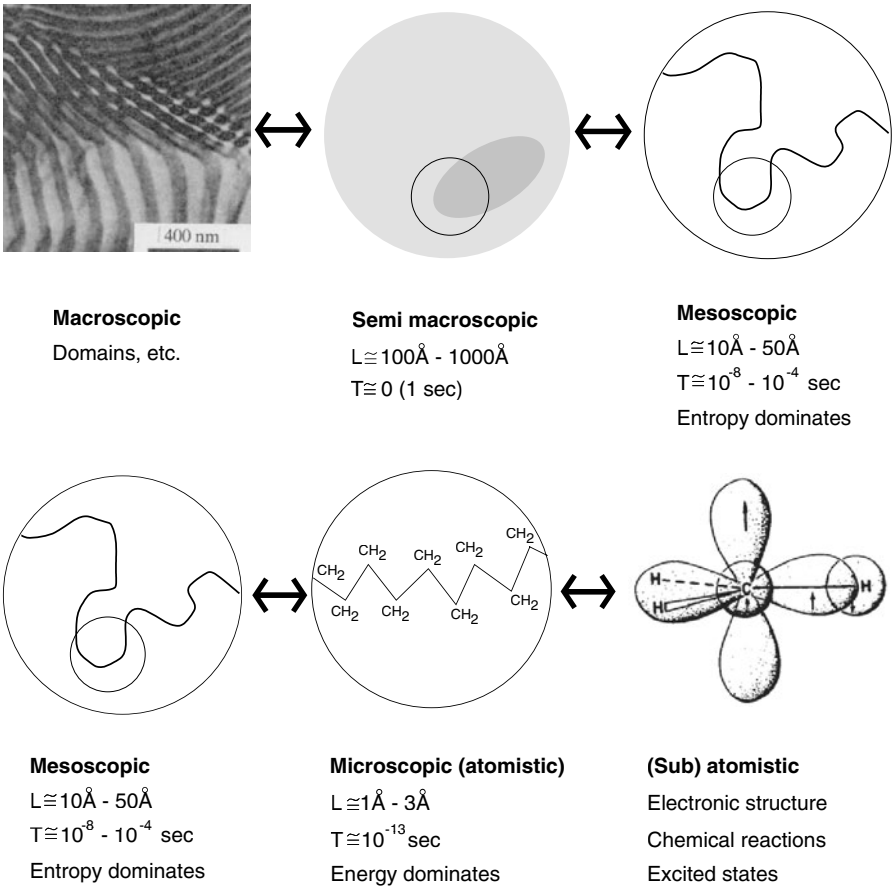


Fig. 1. Polymers exhibit phenomena on many length scales (from entire devices down to electrons) and associated time scales (from years to femtoseconds). Starting from the top left side, one can observe the overall morphology of a polymer material. Looking a little bit more closely, e.g. by marking a chain in a melt or dense solution of otherwise identical chains, the individual polymer can be observed only as a very pale shadow. A typical spatial extension of the shadow is given by the overall coil diameter, as indicated. The characteristic time for this picture to change can vary dramatically depending on chain length and temperature, starting at about $10^{-4}s$ for short chains and 'high temperatures', with essentially no upper limit. Looking again more closely, more of the polymer structure is revealed. This is the universal, entropy dominated coil regime. Again the variation in time can be very large, cf. text. Typical times, as they are present in many experiments, are indicated. Only if the objects are examined much more closely, chemical details of the polymers can be identified. There local chemical details govern the properties and all bond lengths, angles etc. are determined by the energy levels, originating from quantum mechanics. The lower time limit is determined by the highest frequency of oscillations, which depending on the model used are the bond length or bond angle oscillations. To study excited states or reactions, the electronic structure is to be considered and quantum methods are required. Methods for treating individual scales are well developed. The challenge is to connect them systematically.

$\propto N^{3.4}$ for longer chains. Prefactors originating from the microscopic interaction of the monomers, cause an equally large variation of scales. The resulting times can easily reach seconds. If one approaches the glass transition temperature even much longer characteristic times can be observed. Thus a satisfactory numerical description of material properties needs a combination of both aspects.

This illustrates the long standing challenge within the modeling of complex soft matter materials, namely develop methods which allow to cover the range from microscopic to mesoscopic and then on to the (semi-)macroscopic regime. This requires suitable links or mapping schemes. The successful mapping of an atomistic to a mesoscale model (bottom up approach) has the advantage that, on the mesoscopic level, time and length scales are accessible that are far beyond the reach of atomistic simulations, so that qualitatively different physical problems can be treated. A good mesoscale model preserves enough of the original chemical identity of the atomistic model, that it does locally not behave like a generic Gaussian coil, which it has to do on large length scales. It contains just enough information to reproduce certain aspects of, say, polystyrene or polypropylene chains under the corresponding conditions. It is no surprise that in recent years, a number of atomistic-to-mesoscopic mappings have been published; a review encompassing scale bridging from electronic to macroscopic degrees of freedom, cf. Fig.1, is presented in Ref [Bas00]).

Moreover, it has turned out that coarse-grained models are useful not only in their own right to study large scale phenomena. They are also a reliable tool for the generation of well-equilibrated atomistic structures, provided one can perform an inverse mapping from the mesoscopic model back to a fully atomistic model which can then be analyzed. In many cases, the fully atomistic model is needed for comparison with experiment because the experimental information often necessarily involves atoms, like in nuclear-magnetic-resonance (NMR) spectroscopy [Fa00], neutron scattering [Ei99] or positron annihilation spectroscopy [Sch00].

Keeping in mind that a faithful representation of excluded volume is essential in dense liquid simulations, a systematic study of these models has been performed by Abrams and Kremer [Ab01, Ab02]. For these linear chains with a single bead diameter, d_0, and a single prescribed *average* bond length, l_0, as depicted in Fig. 2 was studied. The consequences turned out to be crucial for both static and dynamic

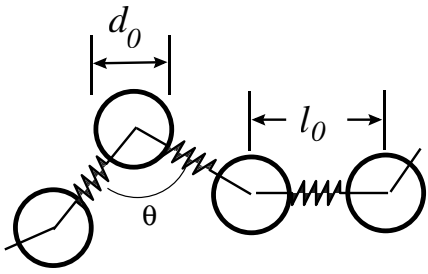


Fig. 2. Schematic of a bead-spring polymer chain. d_0 is the effective bead diameter, l_0 is the average bond length, and θ is a bond angle. From [Ab01].

aspects. We here shortly review results, which were obtained for bulk melts and thin confined melt films, both composed of simple, freely-jointed, bead-spring polymer chains. l_0/d_0 was chosen to vary between 0.73 and 1.34.

The polymer chains are modeled as sequences of N connected identical particles. All particles in the system, except bonded neighbors along a chain, interact according to a purely repulsive shifted 12–6 Lennard-Jones potential, cut off at $r = 2^{1/6}\,\sigma$:

$$U_{LJ}\left(r_{ij}\right) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} + \frac{1}{4}\right] & r_{ij} < 2^{1/6}\sigma; \\ 0 & r_{ij} \geq 2^{1/6}\sigma, \end{cases} \tag{3}$$

with the standard LJ units $[\epsilon] = kT$. All beads have the same excluded volume diameter, d_0, defined as the separation at which the Lennard-Jones potential has a value of $1kT$. With this choice of potential, $d_0 = 1\ \sigma$. Bonds between adjacent neighbors (i,$i+1$) along a chain are enforced by a stiff harmonic potential with a given average bond length, l_0. In most cases the spring is replaced by the LJ-repulsions plus a so called FENE potential [KK96].

The systems were simulated in the NVT ensemble, using a Langevin thermostat, [Gr86] with friction $\Gamma = 0.5$ at constant T = 1.0. For this they ran systems of $M = 80$ chains of length $N = 50$ and a number density of $\rho = 0.85\sigma$.

Here I only consider the influence of the ratio l_0/d_0 on the liquid structure and dynamics in bulk while in Ref. [Ab01, Ab02] also confining surfaces, as they are relevant for the next two contributions of this volume by Delle Site and Abrams are studied. In Fig. 3 the intermolecular radial distribution functions $g(r)$ along with $g^*(r)$ for a corresponding monomer fluid a range of l_0/d_0 ratios is shown. The liquid structure is quite sensitive to this ratio, which holds for both cases whether one considers systems with a fixed volume fraction ϕ or number density ρ. When the bond length is below the excluded volume diameter, monomer-monomer correlations weaken, whereas these correlations are little if at all affected when $l_0/d_0 > 1$. The manner in which chains pack gives rise to the less ordered liquid structure for $l_0/d_0 < 1$, and the similarly ordered liquid structure for $l_0/d_0 > 1$, compared to the simple fluid. Hence, by simply changing l_0/d_0 from about 1.3 to 0.7, one finds a gradual transition from structures determined primarily by monomer packing to structures determined by intramolecular configurational entropy of the chains.

In a similar way the chain diffusion is very sensitive to the choice of l_0/d_0. A commensurate ratio leads to a significantly slowed down diffusion constant, as local cages in the liquid can nicely be created by the system. This maximizes the bead entropy, but at the same time creates significant barriers for diffusion [Ab02].

Fig. 4 illustrates this for a variety of ratios l_0/d_0. Considering that the local excluded volume structure of most simple polymers (e. g. systems without huge side groups) resembles more of a sausage rather than a necklace with well separated pearls, this gives a clear hint towards an optimized modeling of specific polymers on a coarse grained level.

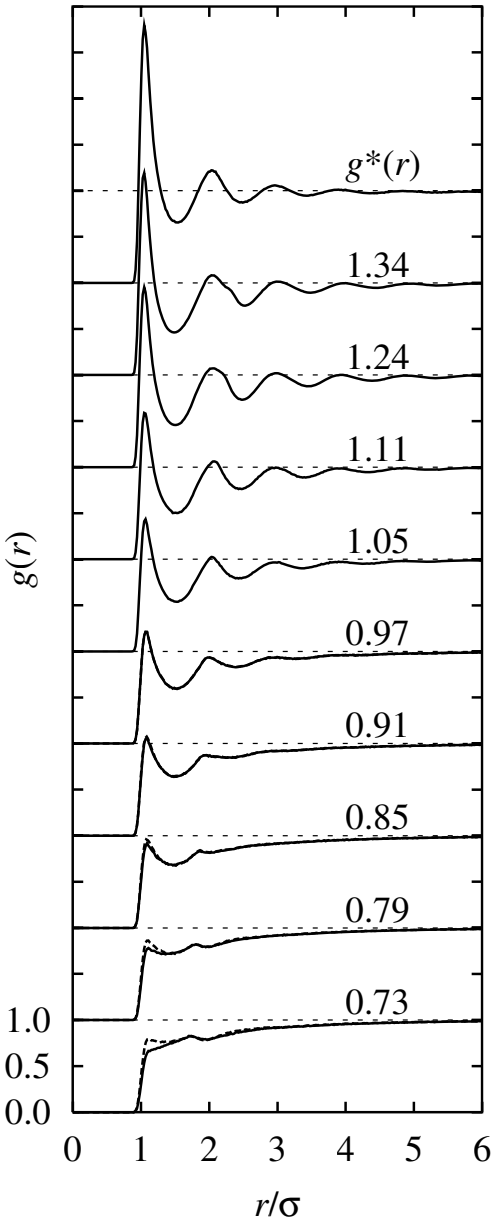


Fig. 3. Intermolecular radial distribution functions, $g(r)$, for bulk bead-spring melts. Solid curves correspond to samples with $\rho = 0.85\ \sigma^{-3}$, and broken curves to those melts for which $l_0/d_0 < 1$ and fixed $\phi = 0.445$. Labels denote bond length, l_0. $g^*(r)$ was computed from a simple repulsive monomeric fluid at $\rho = 0.85\ \sigma^{-3}$. From [Ab01].

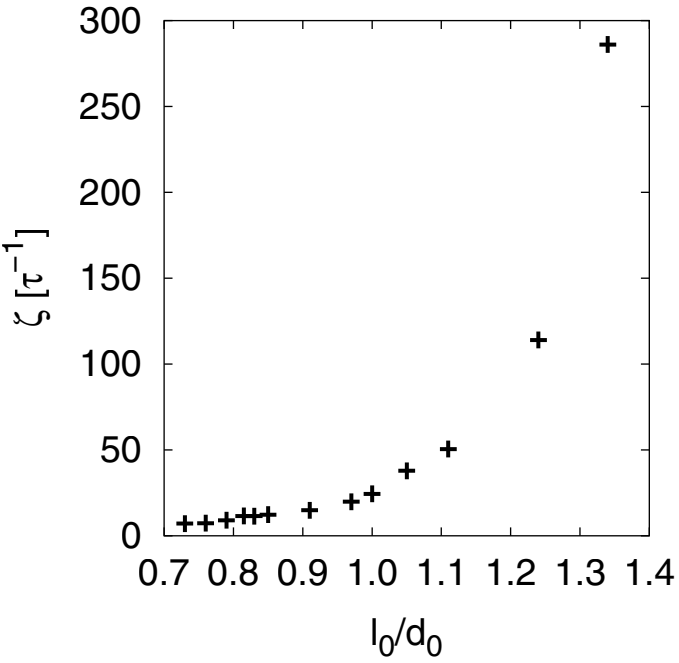


Fig. 4. Monomeric friction constant ξ as a function of l_0/d_0 for constant number density $\rho = 0.85\sigma^{-3}$ [Ab02], τ being the Lennard Jones time unit.

3 Coarse Craining: Atomistic - Mesoscopic

First I give the idea of coarse graining as it was developed for polycarbonate and then will also shortly mention other approaches. More details about specific problems are discussed in the chapters written by L. Delle Site and C. F. Abrams in this volume. The basic original idea was presented in detail in Refs. [Tsc98a] and [Tsc98b]. These papers describe a mapping scheme for bisphenol-derived polycarbonates in which each repeat unit is replaced by two spherical beads; referred to as a "2:1" mapping. Here we focus on a related scheme in which the repeat unit is replaced by four beads, a "4:1" mapping. This is represented schematically in Fig. 5. This much better resembles the aforementioned "sausage" structure than the 2:1 mapping, where the spheres for the benzene rings are replaced by bonds.

The key features of this model are as follows [27, 30]. The bead centers correspond to specific groups of atoms in the molecular structure. Bead centers are mapped to the carbonate carbon, the isopropylidene backbone carbon, and one to each of the phenyl rings centers of mass. With these definitions of the mapping points, the molecular structure of the repeat unit constrains allowable bead-bead distances and bead-bead-bead angles along the coarsened backbone. One important exception is the phenylene-carbonate-phenylene angle, which is constrained to satisfy a *distribution* of angles which reflects an average over the atomic-scale torsional

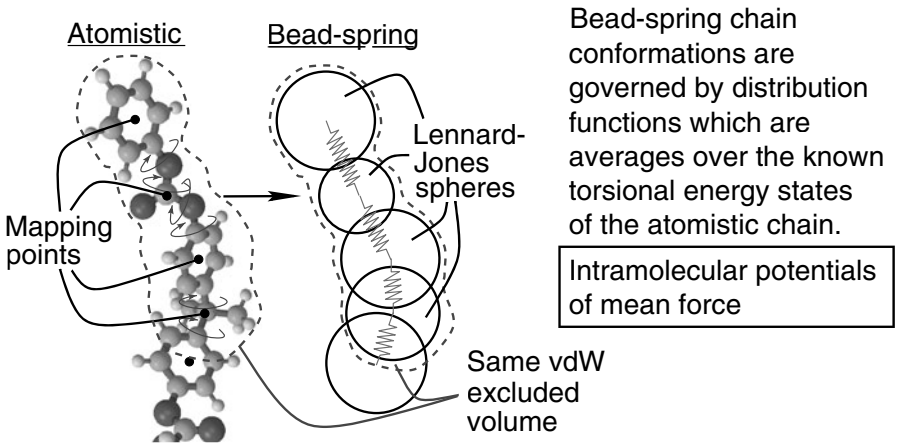


Fig. 5. A schematic representation of the 4:1 mapping scheme for coarse-graining of bisphenol-A-polycarbonate (BPA-PC). A chain-ending repeat unit is shown. From [Ab03]. (See Plate 10 on Page 273)

states "within" this bead. These constraints are translated into intramolecular potentials used directly in molecular dynamics simulations of these coarsened bead-spring objects. The intra molecular potentials for bond lengths, bond angles, and torsions are obtained by a Boltzmann inversion of the corresponding distribution functions, which are sampled from a Monte Carlo simulation of an isolated all atom model. Since this is done for a random walk, these distribution functions can be generated with high precision. The potentials for the microscopic interaction along the back bone are determined by quantum chemical methods. Furthermore, the bead diameters are chosen such that the overall excluded volume of the repeat unit as closely as possible matches that known for BPA-PC [San95]. This prescribes *intermolecular* potentials which are also included in the MD simulations. Of particular interest in this discussion are as before the radial distribution functions, $g(r)$, obtained from bulk simulations. In Fig. 6, we show comonomer-specific $g(r)$'s for the "4:1" and "2:1" liquids. These liquids were composed of chains of ten repeat units each, which corresponds to chains of 23 and 43 coarse grained beads for the 2:1 and 4:1 mapping schemes, respectively. From this data, it is evident that the liquid structure for the different mapping schemes is quite different, as to be expected from the results of the simple bead spring models discussed before. Most notable are the long-ranged decaying oscillations in $g(r)$ for the 2:1 system. This starkly resembles the $g(r)$ of simple dense liquids of spherical particles [Ver67]. In contrast, such long-lived correlations are not apparent in the $g(r)$ from the 4:1 simulations. The differences in the liquid structure are a direct result of the imposed *intramolecular* structure of the two mapping schemes.

The resulting different packing has significant influence on the dynamics as well. It turns out, as expected from [Ab02], that the 4:1 chains move more quickly than the 2:1 chains as a result of the much lower sphere-packing efficiency in the 4:1

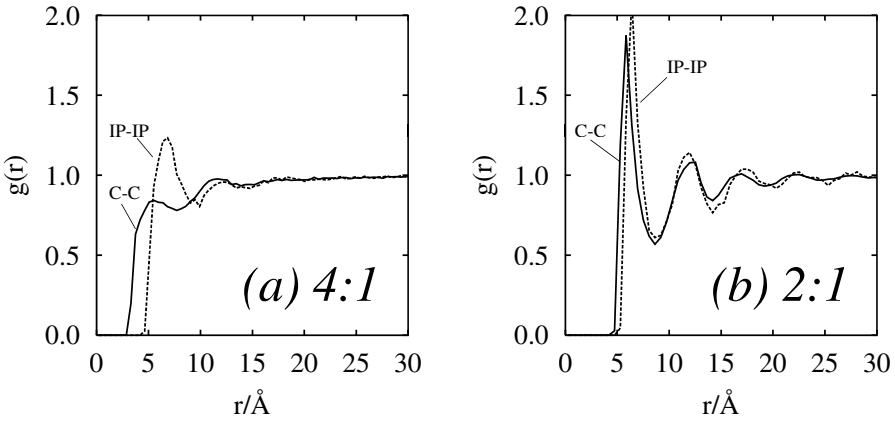


Fig. 6. Carbonate-carbonate ("C-C") and isopropylidene-isopropylidene ("IP-IP") intermolecular radial distribution functions from MD simulations of N=10 BPA-PC: comparison of (a)"4:1" and (b) "2:1" molecular coarse-graining resolutions. From [Ab03].

melts *vis-à-vis* the 2:1 melts. Hence, the bead-specific friction, or resistance to accumulation of mean-squared displacement, is higher in the 2:1 liquids. This slowing down in the bead motion for the 2:1 melt means that more simulation time steps are required to produce the necessary amount of average molecular rearrangement for equilibration, relative to the 4:1 melt. As a result, though more complex, the simulations with the 4:1 mapping scheme use much less computer time than those with the 2:1 mapping scheme. Despite these differences, both models carry the essential microscopic information. The overall end to end distances are the same and when they are used to reconstruct an all atom "sample" the resulting structures from both models are very similar (for $T = 570K$) [Ab03].

4 Other Approaches

As shown above a reliable mesoscopic model of a specific polymer requires rather complex model development. Then however, within the assumptions and approximations made, an almost parameter free mapping can be developed.

A number of different steps, which may vary between approaches have to be taken: (i) The degree of coarse graining, how many real atoms per coarse-grained bead, has to be chosen and the positions of the coarse-grained (CG) beads in relation to the atoms have to be determined. (ii) The form of the intra-chain and inter-chain potentials need to be chosen, if they are not directly derived from distribution. (iii) The free parameters, especially for the nonbonded interactions, have to be optimized in a way that the CG model reproduces the structure of the system. While (i) and (ii) are intellectual challenges, (iii) often is a menial task which should be automated to the extent possible. Recently, Müller-Plathe and coworkers automatically parameterized interaction parameters of mesoscale models for polymers, a method

especially useful in solution. Its purpose is to be able to carry out the parametrization (iii) for a given degree of coarse graining (i) and form of the potential (ii) quickly and reproducibly. First, reference data are obtained, such as structural properties of the polymer of interest. In the present study they concentrated on the sodium salt of poly(acrylic acid) (PAA) as an aqueous solution of about 2 wt.%. They were obtained by performing an atomistic simulation of an oligomer (23 monomers) solvated by about 3200 water molecules. The coarse-grained model contained one bead per monomer centered at the center of mass of the atomistic monomer, reducing the number of polymer atoms by 8. More importantly, the coarse-grained model disposed off the explicit solvent, so that the total number of sites was reduced from approximately 3350 to 23. For a more detailed discussion see [Me00]. The CGA model of the 23-mer of PAA was simulated as an isolated molecule in space, in other words the same oligomer as with the atomistic model. The CG parameters were adjusted, until the target (radial) distribution functions were reproduced satisfactorily. For the optimization, they used a standard amoeba simplex scheme [Pr92]. Note, that every evaluation of involves an entire molecular dynamics (or Brownian dynamics) simulation of the CG system, including equilibration, check for convergence etc. At this point, the apparently straight forward scheme can become technically tricky and computationally expensive [43]. With the CG model parameterized, the simulation was extended to much longer chains of PAA in aqueous solution. The results for the calculated hydrodynamic radius of such chains match well the results from dynamic light scattering are shown in Fig. 7.

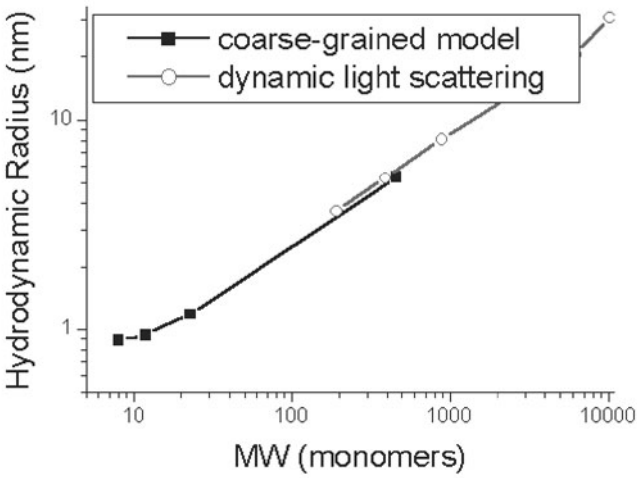


Fig. 7. The hydrodynamic radius of poly(acrylic acid) in aqueous solution as a function of molecular weight. From [Wie02].

This shows that the CG model retains enough of the true identity of PAA to reproduce its structure on a scale much larger than that of the atomistic model from which it was developed. In the meantime this approach has been extended not only to other polymers in solution but also to bulk polymer melts which allow for an all atom simulation of oligomers [Fa99]. Compared to the previous ansatz, this is a more pragmatic way, which in special cases leads more directly to the desired results.

So far, only particle based continuous space approaches were discussed. There are several lattice based methods as well as ideas where a top down ansatz is pursued. There one e. g. starts from the properties of a specific coarse grained model and tries to map typical measurable quantities, such as the end to end distance. In those cases however the one to one correspondence of a model bead and a specific group of atoms often is lost.

5 Outlook

In spite of all progress made over the last years a number of challenges remain. On each level of description, new and improved methods have been developed. Still better models are needed. Most important, however, is the controlled and systematic improvement of links between the different simulation schemes, i.e.:

- Systematic coarse graining procedures, including the inverse mapping step are to be improved and developed. Steps must cover the Micro (many atoms) $<->$ Meso (many monomers) $<->$ Macro (many chains) regimes and link to quantum simulations at the low end and to self consistent field calculations and finite element like approaches at the upper end.

In order to achieve this and to predict macroscopic material properties from first principles, much effort is needed over the coming years. In particular, a few key ingredients have to be developed or improved, such as

- quantum simulations of reasonably sized systems (Car-Parinello techniques, path integral quantum Monte Carlo, combinations of both): coupling electronic and conformational degrees of freedom (beyond the ansatz of [De02])
- improved methods to parameterize and validate force fields for atomistic classical molecular simulations, especially suitably parameterized inter-molecular interactions
- new methods for static and dynamic studies on the semimacroscopic to macroscopic level, such as dissipative particle dynamics for composite materials based on the microstructure of the polymers.
- adaptive multiscale schemes, which allow to systematically vary the resolution of the simulations locally and temporarily

Acknowledgement The work described here is supported by the BMBF Kompetenzzentrum Materialsimulation (Förd. Kennz. 03N6015). I would like to thank the many coworkers (especially F. Müller-Plathe, L. Delle Site, and C. F. Abrams) and

students who were involved in various aspects of the described examples over the past years.

References

[KK96] K.Kremer in [Bi96]

[Bas00] J.Baschnagel, K.Binber, P.Doruker, A.A.Gusev, O.Hahn, K.Kremer, W.L.Mattice, F. Müller-Plathe, M.Murat, W.Paul, S.Santos, U.W.Suter and V.Tries: Adv.Polym.Sci. **152**, 41 (2000).

[Bi95] K.Binder (ed.): *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*. Oxford Univ.Press, Oxford, 1995

[Mo94] L.Monnerie and U.W.Suter (eds.):*Atomistic Modelling of Physical Properties*, in Advances in Polymer Science **116**, Springer, Heidelberg (1994).

[Pe02] V. Percec, M. Glodde, T. K. Bera, Y. Miura, I. Shiyanovskaya, K. D. Singer, V. S. K. Balagurusamy, P. A. Heiney, I. Schnell, A. Rapp, H. W. Speiss, S: D. Hudson, H. Duan: Nature *419*, 384 (2002)

[Ge79] P.G.de Gennes: *Scaling Concept in Polymer Physics*. Cornell University Press, Ithaca, New York, 1979.

[Do86] M.Doi and S.F.Edwards:*The Theory of Polymer Dynamics*. Oxford Scientific Publications, Oxford, 1986.

[Bi96] K.Binder and G.Ciccotti (eds.): *Monte Carlo and Molecular Dynamics of Condensed Matter Systems*. Oxford University Press, Oxford, 1996.

[Ma97] R.Martonak, W.Paul and K.Binder: J.Chem.Phys. **106**, 8918 (1997)

[Ge92] M.D.Gehlsen, J.Rosendale, F.S.Bates, G.D.Wingall, L.Hansen, K.Almdal: Phys.Rev.Lett **68**, 2452 (1992).

[Fa00] R.Faller, F. Müller-Plathe and A.Heuer: Macromolecules **33**, 6602 (2000)

[Ei99] J.Eilhard, A.Zirkel, W.Tschöp, O.Hahn, K.Kremer, O.Schärpf, D.Richter and U.Buchenau: J.Chem.Phys. **110**, 1819 (1999)

[Sch00] H.Schmitz and F. Müller-Plathe: J.Chem.Phys **112**, 1040 (2000).

[All87] M.P.Allen and D.J.Tildeslesy, *Computer Simulations of Liquids*. Oxford University Press, New York, 1987.

[Gr86] G.S.Grest and K.Kremer: Phys.Rev.A **33**, 3628, (1986).

[Kre90] K.Kremer and G.S.Grest: J.Chem.Phys. **92**, 5057 (1990).

[Ba99] S.Barsky: J.Chem.Phys. **112**, 3450 (1999).

[Du93] E.R.During, K.Kremer and G.S.Grest: Macromolecules **26**, 3421 (1993).

[Du94] E.R.During, K.Kremer and G.S.Grest: J.Chem.Phys **101**, 8169 (1994).

[Pu00] M.Pütz, K.Kremer and R.Everaers: Phys.Rev.Lett **84**, 298 (2000).

[Ay00] E.M.Aydt and R.Hentschke: J.Chem.Phys. **112**, 5480 (2000).

[Mu89] M.Murat and G.S.Grest: Macromolecules **27**, 418 (1989).

[Gr94] G.S.Grest: Macromolecules **22**, 418 (1994).

[Cs00] F.S.Csajka and C.Seidel: Macromolecules **33**, 2728 (2000).

[Ab01] C.F.Abrams and K.Kremer: J.Chem.Phys. **115**, 2776 (2001).

[Ab02] C.F.Abrams and K.Kremer: J.Chem.Phys. **116**, 3162 (2002).

[Bi90] I.Bitsanis and G.Hadziioannou: J.Chem.Phys. **92**, 3827 (1990).

[Tsc98a] W.Tschöp, K.Kremer, J.Batoulis, T.Bürger and O.Hahn: Acta Polym. **49**, 61 (1998).

[Tsc98b] W.Tschöp, K.Kremer, J.Batoulis, T.Bürger and O.Hahn: Acta Polym. **49**, 75 (1998).
[Ab03] C.F.Abrams and K.Kremer: Macromolecules, **36**, 260 (2003).
[De02] L. Delle Site, C. Abrams, A. Alavi, K. Kremer: Phys. Rev. Lett. **89**, 156103 (2002).
[San95] I.C.Sanchez and J.Cho: Polymer **36**, 2929 (1995).
[Ver67] L.Verlet: Phys.Rev. **159**, 98 (1967).
[Re01] D.Reith, H.Meyer and F. Müller-Plathe: Macromol. **34**, 2235 (2001).
[Me00] H.Meyer, O.Biermann, R.Faller, D.Reith and F. Müller-Plathe: J.Chem.Phys. **113**, 6264 (2000)
[Pr92] W.H.Press, S.A.Teukolsky, W.T.Vetterling and B.P.Flannery: *Numerical Recipes in C* (2nd Edition, Cambridge University Press, New York (1992).
[Fa99] R.Faller, H.Schmitz, O.Biermann and F. Müller-Plathe: J.Comp.Chem. **20**, 1009 (1999).
[Wie02] D.Reith, B.Müller, F. Müller-Plathe and S.Wiegand: J.Chem.Phys. **116**, 9100 (2002).

Polymers near a Surface: An *ab initio* Density Functional based Multiscale Modeling Approach

Luigi Delle Site[1]

Max-Planck-Institute for Polymer Research, D-55128 Mainz, Germany
dellsite@mpip-mainz.mpg.de

1 Introduction

The study of surface-polymer interface properties (e.g. adhesion behaviour and its influence on the melt morphology) is a field of large interest for scientific as well as technological reasons. The theoretical description and prediction of such properties plays a crucial role in improving technological processes but at the same time requires the development of accurate tools of study. Current analytical and semi-analytical approaches (see for example Refs.[Ei94, Fl98]) describe polymers and surfaces as ideal mathematical objects while simulation studies have been restricted to bead-spring models of polymers interacting with a purely repulsive wall or a wall with an arbitrary attraction (see for example Ref.[Ab02a]). These approaches, certainly useful for a general description, cannot take into account local polymer-surface interactions which depend on the specific chemical structure of the system's components and occur at the electronic level. This means that, for a realistic description, the crucial question is whether or not selective adsorption can influence global conformations or, in other terms, what is the effect of the interplay between energy (of selective adsorption) and entropy (of polymer conformations). Currently, a full quantum mechanical treatment of such systems is not possible, thus a multiscale modeling approach, where parameters derived from quantum-mechanical calculations are plugged into meso- and macroscopic simulation models, would represent an adequate and feasible tool of study. In this case, the idea behind the multiscale approach is to describe the adsorption process at the molecular level by employing an *ab initio* Density Functional technique and to use its results to properly parameterize a bead-spring coarse-grained model for simulating large systems. In the next sections we present some relevant applications of the multiscale approach.

2 Adsorption of Bisphenol-Polycarbonate-*A* on a Ni(111) surface.

We have chosen Bisphenol-A-Polycarbonate (BPA-PC) (see Fig.1(a)) on a Ni(111) surface as a test system for the implementation of our multiscale approach [De02,

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

Ab02b, Ab03]. This system is of high interest for technological processes such as the production of optical data storage devices and, at the same time, BPA-PC submolecules are common to many polymers. Since *ab initio* Density Functional calculations are computationally rather demanding, our strategy should consist of combining feasibility, accuracy and efficiency. For this reason the first step of our approach consists in cutting the BPA-PC chain into comonomeric submolecules which are small enough to study the interaction of each with the nickel surface. This procedure on one hand reduces the computational load and on the other allows the modeling of other polymers whose subunits are those studied in this case. However this implies that the role of the hydrogens, which are used to saturate the free valences of the polymer fragments, is negligible in the adsorption process of the molecule. Test calculations on large molecules which can be divided into submolecule show that indeed this approximation works rather well [De03a] for metal surfaces, of course one does not expect the approximation to be valid for surfaces which strongly interacts with hydrogens (e.g. oxide surfaces).

The subunits of BPA-PC considered are: carbonic acid representing carbonate, propane representing isopropylidene and benzene or phenol representing phenylene and phenoxy chain ends (see Fig.1(b)).

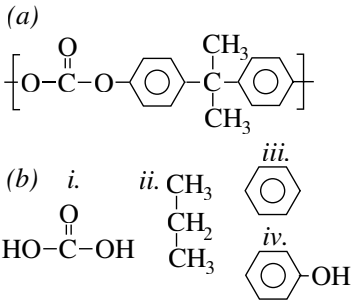


Fig. 1. (a) Schematic representation of the chemical structure of the BPA-PC repeat unit. (b) Subunit's division of the BPA-PC repeat chemical unit. Carbonic acid (i) corresponds to the carbonate group, propane (ii) to isopropylidene and phenol (iii) or benzene (iv) to the phenylene.

2.1 *Ab initio* calculations.

The *ab initio* molecule-surface study, as a part of a multiscale description, is not the standard one but deserves particular attention. The crucial requirement is that these submolecules, together, should mimic the behaviour of a monomer of a BPA-PC chain near the surface. In order to fulfill such a requirement, we select molecular orientations with respect to the surface which are compatible with chain conformations at the surface. Each selected orientation is then taken as a starting configuration for a molecule-surface geometry optimization and final optimized geometries are checked

to be still compatible with chain conformations at the surface. The resulting adsorption energy indicates the strength of the molecule-surface interaction. By following this scheme we studied the interaction of carbonic acid, propane, and phenol (benzene) at each high symmetry site of a Ni(111) surface. We used the CPMD code [CPM] in the FEMD version of A.Alavi [Al94, Al96]. The orbital cut-off was set to 60 Ry. We used the PBE[Pe96] generalized gradient approximation (GGA). The surface is represented by four close-packed layers of Ni $\{111\}$ (lattice parameter a_0 = 3.543 Å), with the top two layers allowed to relax. We used a (2×2) lateral supercell for carbonic acid and propane adsorption, and a (3×3) cell for benzene and phenol, employing $4 \times 4 \times 1$ and $3 \times 3 \times 1$ k-point mesh for the smaller and larger cells, respectively. Further technical details of the calculations are given in Refs.[De02, Ab02b, Ab03, De03b].

2.2 *Ab initio* Results and Modeling.

Fig. 2 shows the adsorption energy of carbonic acid as a function of the central carbon distance from the surface. The curve was obtained by taking the energetically most stable configuration, obtained from the geometry optimization of the various selected molecular configurations at each surface high symmetry site, and adiabatically moving it towards (or away from) the surface in steps. At each step the z-coordinate of the central carbon was kept fixed and the remaining coordinates were relaxed yielding a z-dependent adsorption energy. These results show that the

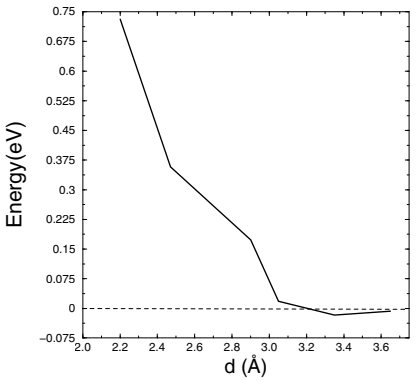


Fig. 2. Adsorption energy of carbonic acid as a function of the central carbon distance from the surface.

molecule experiences strong increasing repulsion below 3.2 Å, at this distance it reaches an adsorption energy of about 0.01 eV which is both negligible compared to the inherent error of the *ab intio* calculations (≈ 0.05 eV) and smaller than the characteristic thermal energies in typical melt processing of polycarbonates ($E \approx k$T, with T=$570K$, kT$\approx$ 0.05 eV). For this reason we can model the carbonate bead of

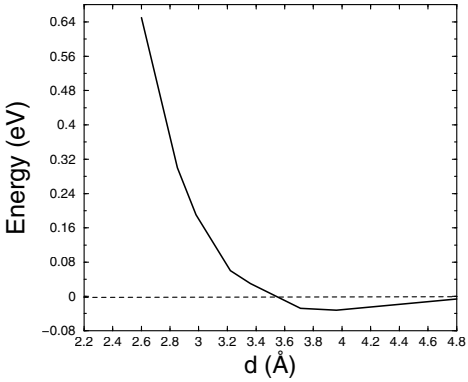


Fig. 3. Same plot as the previous figure for propane.

the polymer as a sphere interacting with a uniform repulsive wall. Fig. 3 is the equivalent of Fig. 2 for propane. The qualitative and quantitative behaviour is very similar to that of the carbonic acid and as a consequence the isopropylidene bead is modeled as the carbonate bead. Table 1 reports the adsorption energies of phenol and benzene at the high symmetry sites of the surface [De03b, Mi01]. For benzene results were available in literature thus we repeated only the calculation at the most energetic site in order to reproduce available results as a check of validity of our technical setup; our results agree with those of previous calculations and experiments [Mi01, Ya01]. Adsorption energies for benzene are qualitatively and, within a 0.1-0.2 eV difference, quantitatively equivalent to those of phenol. All the values of Table 1 correspond to

Table 1. Adsorption energies for phenol and benzene at the four high symmetry sites of the surface. For benzene the values were taken from Ref[13].

$E_{ad}(eV)$	Phenol	Benzene
FCC	0.79	0.94
HCP	0.84	0.91
atop	0.20	0.42
bridge	0.91	1.00/(1.05 our value)

a geometry where the carbon ring lays parallel to the surface at an average distance of about 2.0 Å. Differently from the previous two molecules, except for the atop site, the adsorption energy is relatively high. Since the atop site is statistically not relevant and the diffusion barrier small, one can conclude that in general phenol and benzene strongly bind to the nickel surface. However, as Fig. 4 and Fig. 5 show, such an interaction is short-ranged as a function of the distance (it decays to 0.03 eV beyond 3.0 Å) as well as a sensitive function of the inclination of the carbon ring (it decays to 0.05 eV beyond 40°).

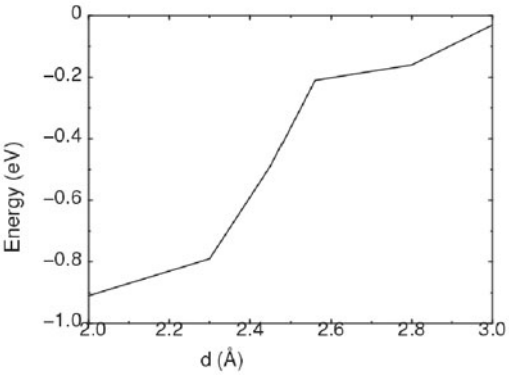


Fig. 4. Adsorption energy of phenol at the bridge site as a function of the distance from the surface.

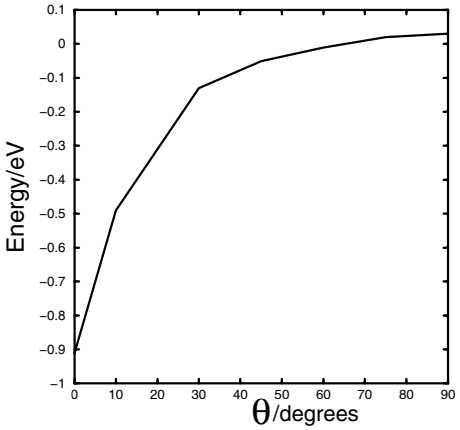


Fig. 5. Adsorption energy of phenol at the bridge site as a function of the inclination of the carbon ring with respect to the surface.

For this reason, although phenol (benzene) experiences strong adsorption energy in isolation, due to the short range nature of this interaction, it is sterically hindered by the two adjacent submolecules to adsorb when incorporated into the BPA-PC; this is pictorially illustrated in Fig. 6. For this reason internal phenylene beads are modeled as the carbonate and isopropylidene beads. However phenoxy chain-ends are not in this way sterically hindered and can strongly adsorb at the surface as illustrated in Fig. 7. As a consequence beads corresponding to phenoxy chain-ends are modeled as spheres which experience a strong short-ranged attraction near the surface.

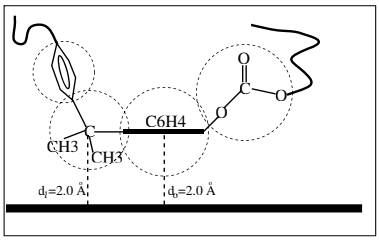


Fig. 6. Schematic representation of an isopropylidene-phenylene-carbonate sequence in a BPA-PC chain at the surface. The adsorption configuration of an internal phenylene leads to a distance isopropylidene-surface (d_1) for which the total repulsion is much larger than the total attraction.

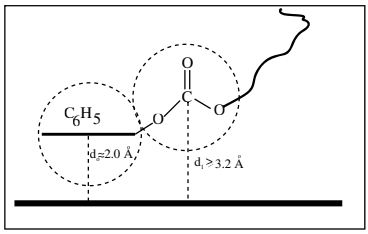


Fig. 7. Schematic representation of a (phenoxy) chain-end-carbonate sequence in a BPA-PC chain at the surface. The chain end is not sterically hindered to adsorb at the surface since the adjacent carbonate is at a distance (d_1), where it does not experience repulsion.

One should note that dispersion effects, which are not reproduced by current density functionals, were neglected. This approximation is justified since the relative and not the absolute adsorption energies are those required for developing a valid coarse-grained model. In this case the separation between the energy scales relative to the different submolecules is at least one order of magnitude larger than the expected dispersion effects. On the other hand, dispersion might have an effect onto the sensitivity of the angle dependence adsorption energy of the benzene. However, as above, this aspect does not play an important role for the modeling procedure. In fact, the angle dependence study simply shows that as soon as the benzene is not in its optimal orientation with respect to the surface the strength of the interaction rapidly decays; at the same time it is not likely that dispersion effects will dramatically change this qualitative behaviour. This is the relevant information required for the modeling procedure. The final coarse-grained model is then sufficiently accurate to describe general interface properties such as the distribution of the different bead types at the surface or the influence of a specific bead adsorption onto the global melt morphology. Of course a valid estimate of dispersion effects would be necessary for accurate quantitative predictions.

3 Coarse-grained Simulation: Results

The *ab initio* modeling of the previous sections was incorporated into a bead-spring coarse-grained model (see Fig. 8) and simulations of BPA-PC melts for chain lengths of 10 and 20 repeat units were carried out. Details of the inter and intra-chain poten-

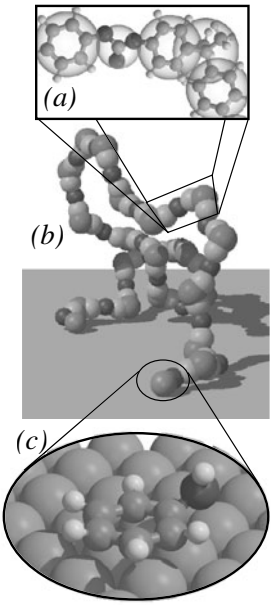


Fig. 8. Pictorial illustration of the multiscale model for BPA-PC: (a) coarse-grained beads (transparent spheres) superimposed on the underlying chemical structure; (b) Coarse-grained model of an $N = 20$ BPA-PC molecule, with ends adsorbed on the surface. (c) A phenol molecule adsorbed on the bridge site of a Ni (111) surface; configuration computed via CPMD simulation. (From Ref.[4]) (See Plate 11 on Page 274)

tial parameterizations together with more simulation details and results are given in a contribution to this issue by C.F.Abrams. Here, we report only results which are relevant to characterize the utility of our procedure. The main results of the bead-spring coarse-grained simulation are that chain ends are strongly localized at the surface and a significant layering in the density of beads is observed; the chain end strong adsorption determines not only properties at the interface but even at larger distances from the surface. In this way we can answer our initial question about the effect of selective adsorptions on global conformations; we have characterized the interface properties as energy dominated. Moreover, even at large distances from the surface, properties are influenced by the interplay between energy and entropy.

4 Conclusions and Outlook

We presented a multiscale modeling approach to simulate polymers on a surface. In this approach, an *ab initio* Density Functional study of molecular adsorption is employed to construct an appropriate bead-spring coarse-graining model which incorporates the microscopic results and predicts macroscopic properties. We have shown an application to BPA-PC interacting with a Ni(111) surface. This procedure has been extended to study the effect of different chain ends for BPA-PC on a Ni(111) [De03c] and to the study of BPA-PC on a Ni(221) step surface, where we found that internal phenylenes can adsorb inducing a certain degree of nucleation at the step [De03d, De03e]. By using the calculations for benzene, we could extend this modeling procedure to the study of amino acids, containing aromatic submolecules, on a Ni(111) and Au(111) surface [De03a]. In particular, for the case of phenylalanine, we directly checked that our prediction based on the submolecules modeling approach was in good agreement with results obtained from calculations where the full system was considered. Work in progress involves the study of thiol and thiolate terminating amino acids interacting with Ni(111) and Au(111); water on Ni(111) is also being studied as a part of a more general multiscale approach where the final goal is to investigate amino acids in solution in contact with a metal surface. As we have shown for BPA-PC, this modeling procedure can predict interface properties which cannot be described by other theoretical procedures; at the same time allows the study of systems which could not be accurately considered so far, as the various extensions and the work in progress show. [De03d, De03e]

5 Acknowledgments

I am grateful to C.F.Abrams, K.Kremer, M.Parrinello, A.Alavi and R.O.Jones for many helpful comments and suggestions. The work was supported by the Bundesministerium für Bildung and Forschung (the German Federal Ministry of Education and Research), grant No.03 N 6015 on materials simulations.

References

[Ei94] E.Eisenriegler: Polymers Near Surfaces. World Scientific (1994).

[Fl98] G.J.Fleer, M.A.Cohen Stuart, J.M.H.M.Scheutjens, T.Cosgrove and B.Vincent:Polymers at Interfaces. Chapman and Hall (1998).

[Ab02a] 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**, 3162-3165 (2002).

[De02] 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).

[Ab02b] C.F.Abrams, L.Delle Site and K.Kremer: Multiscale Computer Simulations for Polymeric Materials in Bulk and near Surfaces In: Bridging time scales: Molecular Simulations for the next decade. P.Nielaba, M.Mareschal, G.Ciccotti Eds. Lecture Notes in Physics 605 pg.143 ff. Springer-Verlag (2002).

[Ab03] C.F.Abrams, L.Delle Site and K.Kremer: Dual-Resolution Coarse-Grained/Atomistic Simulation of the Bisphenol-A-Polycarbonate/Nickel Interface. Phys.Rev E, **67**, 021807 (2003).

[De03a] L.Delle Site and K.Kremer: Multiscale modeling of polymers on a surface: From ab initio Density Functional calculations of molecular adsorption to large scale properties. IJQC to be published.

[CPM] CPMD v. 3.4.1, J.Hutter, A.Alavi, T.Deutsch, M.Bernasconi, S.Goedecker, D.Marx, M.Tuckerman, M.Parrinello. Max-Planck-Institut für Festkörperforschung and IBM Zurich Research Laboratory, 1995-1999.

[Al94] A.Alavi, J.Kohanoff, M.Parrinello, and D.Frenkel: Ab-intio molecular-dynamics with excited electrons. Phys. Rev. Lett., **73**, 2599-2602 (1994).

[Al96] A.Alavi: Path integrals and ab initio molecular dynamics In: Monte Carlo and Molecular Dynamics of Condensed Matter Systems, K.Binder and G.Ciccotti (Eds.). Italian Physical Society, Bologna (1996).

[Pe96] J.P.Perdew, K.Burke and M.Ernzerhof: Generalized gradient approximation made simple. Phys. Rev. Lett., 7**73**, 3865-3868 (1996).

[De03b] L.Delle Site, A.Alavi and C.F.Abrams: Adsorption energies and geometries of phenol on the (111) surface of nickel: An ab initio study. Phys.Rev.B, **67**, 193406 (2003).

[Mi01] F.Mittendorfer and J.Hafner: Density-functional study of the adsorption of benzene on the (111), (100) and (110) surfaces of nickel. Surf.Sci., **472** 133-153 (2001).

[Ya01] S.Yamagishi, S.J.Jenkins and D.A.King: Symmetry and site selectivity in molecular chemisorption: Benzene on Ni111. J.Chem.Phys.,**114** 5765-5773 (2001).

[De03c] 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. Submitted for publication.

[De03d] L.Delle Site and D.Sebastiani: Adsorption of Benzene on a (221) Surface of Nickel: An ab initio Density Functional study. Submitted for publication.

[De03e] L.Delle Site, S.Leon and K.Kremer: Specific interaction of polymers with surface defects: Example of polycarbonate on nickel. Submitted for publication.

Dual Resolution Molecular Simulation of Bisphenol-*A* Polycarbonate Adsorption onto Nickel (111): Chain Length Effects

Cameron F. Abrams

Department of Chemical Engineering, Drexel University, Philadelphia, Pennsylvania 19104, U.S.A., cfa22@drexel.edu

Summary. We discuss results of molecular dynamics simulations of dense liquids of BPA-PC adjacent to (111) nickel surfaces. The BPA-PC molecule is modeled using a dual-resolution coarse-grained representation, in which the chemical repeat unit is represented by four spherical beads, each roughly corresponding to a comonomeric functional group, except at the chain ends, where the terminal carbonate groups are represented atomistically. This dual resolution scheme is necessary to give access to an orientational degree of freedom upon which the polymer/surface interaction potential sensitively depends. The results expand upon those of Ref. [Abr03b], in which chains of length N = 10 repeat units were considered, by considering chains of length N = 20. We observe that the structure of the liquid near the wall is sensitively affected by the strong attraction of the chain ends to the surface for both chain lengths. The liquid forms two layers: in the innermost layer near the wall, most chains have both ends adsorbed, while in the outermost layer, most chains have a single end adsorbed. This structure leads to an interesting profile in chain orientation, where chains are flattened in the innermost layer and stretched in the outermost layer. The overlap between these two layers is more diffuse in the case of the N = 20 chains.

1 Introduction

Simulation of specific polymer systems has recently been greatly advanced by employing systematically derived coarse-grained molecular models [Bas00, Paul91, Kre01, M-P02]. These models vary in the details of their development and implementation, yet all seek to construct particle- or lattice-based chain-like objects whose constituents represent at most a few chemical repeating units of any specific polymer. This is meant to circumvent the great difficulty in producing equilibrated samples of atomically-resolved polymers due to the $O(10)$ orders of magnitude spread between the resolution required in time for standard atomistic molecular dynamics (MD) simulation (10^{-15} s) and the slow molecular relaxation times ($\sim 10^{-5}$ s) which must be achieved. Monte Carlo simulation fares no better as an alternative approach, especially in dense muiltimolecular systems, because local moves are constrained to impractically small values by steep bonded potentials, and enormous numbers of

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

successful moves have to accumulate in order to move chain sections larger than a repeat unit or so. Integrating out these fast motions, or put another way, averaging over their underlying steep potentials, allows us in principle to construct models that can produce equilibrated configurations with relatively much less computational effort.

Coarse-graining aims to guarantee that the chain conformations in a simulation sample represent true equilibrium conformations of the specific polymer considered. A central feature of coarse-graining is that the models retain only as much unique and relevant information as needed about the specific polymer(s) under investigation, using significantly fewer degrees of freedom (*i.e.*, particles) than required for full atomistic detail. The coarsened degrees of freedom must be constrained within ensembles of configurations which represent an appropriate average over the microscopic atomic-scale potential energy surface for the fully resolved system. Once an equilibrated sample at the coarse-grained level is generated using an appropriate simulation technique which samples these ensembles, atomic details can be inverse mapped to study atomic-scale properties and processes on then appropriately short length and time scales [Tsc98b].

Our focus in recent years has been developing a method for coarse graining melts of bisphenol A-polycarbonate (BPA-PC) [Tsc98a, Tsc98b, Hahn01, Del02, Abr03a] (Fig. 1). BPA-PC not only provides a challenging test case for this type of modeling, it is also by far the most utilized and intensively studied variety of polycarbonate, thanks to its many valuable material properties, such as high impact strength, ductility, glass transition and melting temperatures [Mor88]. Coarse-grained simulations can potentially fill a gap that now exists between traditional atomistic simulation and experiments, by addressing longer length and time scales, allowing one to test hypotheses regarding inter- and intramolecular interplay in packing and entanglement that are otherwise more difficult to approach.

Metal surfaces are particularly relevant for industrial processing of polycarbonate, and indeed many other synthetic polymers. The interplay between specific local adsorption of organic chain segments and the much larger scale chain conformations is difficult to capture with traditionally simple molecular and atomic-scale models. Our contributions so far have presented a technique involving appropriate coupling of information from detailed *ab initio* calculations of small molecule/surface interactions into coarse-grained molecular dynamics simulations of liquids of comparatively much larger macromolecules [Del02, Abr03b]. For the purposes of these proceedings, we are particularly interested in understanding the adhesion behavior of BPA-PC liquids next to crystalline nickel surfaces, and how chain length affects this behavior.

Polymer/surface interactions are especially challenging to capture with coarse-grained models due to the specificity of the interactions for various comonomeric groups in complicated polymers, such as BPA-PC. The challenge lies in the fact that the averaging over the microscopic potential energy surface which gives coarse-grained potentials for bulk BPA-PC is strictly not valid when chains strongly interact with surface, which is the case for the BPA-PC/Ni. For this reason, we have constructed a dual-resolution representation of BPA-PC– combining both coarse-grained and atomic-scale resolution in the same representation – in order to capture as much

of the specificity of the polymer/surface interaction as possible, while maintaining the computational advantages of coarse-graining [Abr03b]. In a recent publication, we presented the details of this dual-resolution scheme (which are briefly recapitulated below) and presented preliminary results on the structure of a 10-repeat-unit BPA-PC liquid near a Ni (111) surface [Abr03b]. Here, we expand upon the results presented in Ref. [Abr03b] to consider the effects of longer chains on the structure of the melt near the surface.

2 The Dual-Resolution Coarse-Grained Model of BPA-PC

2.1 The 4:1 Single-Resolution Model

For a complete presentation of the dual-resolution model of BPA-PC and simulation techniques, we refer the reader to Ref. [Abr03b]. Fig. 1 shows the chemical structure of a BPA-PC molecule composed of n repeat units. The repeat unit is constructed of four sequential subunits: a phenylene, an isopropylidene, a second phenylene, and a carbonate. The ends of a BPA-PC molecule are phenyls connected to carbonates. Because the terminal phenyls are each bound to an oxygen of their respective carbonate, we term these "phenoxy" ends. Other end groups apart from phenoxies are currently being considered in work that is not discussed here.

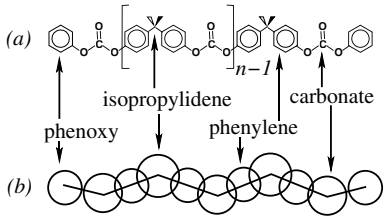


Fig. 1. (*a*) Schematic representation of the structure of bisphenol A-polycarbonate. The relevant comonomeric groups are labeled. The nth repeat unit is outside the brackets because we distinguish *terminal* carbonate groups from *internal* carbonate groups in the dual-resolution model. (*b*) Schematic representation of the 4:1 coarse-grained model of BPA-PC, detailed in Ref. [Abr03a].

The 4:1 coarse-grained representation of BPA-PC [Abr03a] replaces the atomic structure of the repeat unit, which contains 33 atoms, with four spherical beads, two of which are centered on the backbone carbons of the carbonate and isopropylidene groups, respectively. The other two beads "float" at a fixed distance along a line connecting adjacent pairs of carbonate-isopropylidine ("c"–"i"), and are meant to roughly occupy space ascribed to the phenylene ("p")groups. The diameters of these beads are chosen so that they have proper relative sizes and as a group occupy the proper van der Waals excluded volume of the repeat unit. These diameters parameterize repulsive pairwise Weeks-Chandler-Andersen 12–6 potentials [Wee71] for the non-bonded interactions among beads.

The beads adjacent along the chain are connected by stiff harmonic bonds whose lengths are determined from the atomic geometry. In addition to this connectivity, and the non-bonded pair potentials, the overall potential also operates on three types of bond angles: c-p-i, c-i-c, and i-c-i. All c-p-i angles are held at 180° by stiff harmonic potentials. The angles defined by c-i-c and i-c-i triples (skipping the bridging phenylene beads) are constrained according to Boltzmann-inverted potentials tabulated via Monte-Carlo averaging over the atomic scale potential. At a specified temperature, the value of any c-i-c or i-c-i angle is thus constrained to the correct distribution consistent with the atomic scale potential.

The philosphy behind this method of coarse-graining has been discussed previously [Tsc98a, Abr03a]. Briefly, there are three major assumptions used. (1) The coarse-grained degrees of freedom are uncorrelated, allowing us to write a coarse-grained potential with additive contributions from each degree of freedom. (2) The intramolecular potential at the atomic level is not influenced by intermolecular degrees of freedom. This is a safe assumption for most synthetic (carbon-based) polymers at temperatures below those at which thermal decomposition can occur. (3) The intermolecular interactions at the coarse-grained level are approximated to an acceptable level of accuracy by pair-wise repulsive potentials, at least for simulating equilibrium structures.

2.2 Adapting the 4:1 Model for Surface Studies: *Ab Initio* Calculations

The coarse-graining procedure described above yields models which are most appropriately used in bulk simulations. However, it is of interest presently to understand how polymer molecules behave at metal interfaces. We have chosen to consider the specific system of BPA-PC on nickel (111), for the reasons alluded to in the Introduction. What changes must be made to the current 4:1 model, if any, to allow us to use it to study the BPA-PC-Ni interface at the molecular level?

To answer this question, it would be desirable to perform *ab initio* calculations of at least a single BPA-PC molecule in the vicinity of a semi-infinite slab of nickel atoms with a (111) surface arrangement. Were such a simulation tractable, we could use it to determine the exact nature of any specific interactions between any grouping of atoms in the chain and any grouping of atoms on the surface. Appropriately averaging over all possible interaction geometries would then yield probability distributions which could be inverted into potentials for use in coarse-grained simulation.

There are two problems here. The first is that the *ab initio* calculations described are impossible based purely on the number of atoms required, with current or even next-generation computing power. The second is, even if they were possible, and even if we could gain enough statistical strength to perform the averages appropriately, we would have to greatly expand the parameter space to consider how specific interactions with the surface alter the *intramolecular* potential of the molecule. Probing all possible conformations and orientations of a chain of even one repeat unit next to a sufficently large slab of nickel using *ab initio* techniques would be a monumental effort.

Our strategy has been to consider first small molecules which correspond to the chemical subunits of BPA-PC [Del02, Abr03b]. This is motivated not only by the need of computational tractability, but also has the important advantage that the subunits we study are also important for polymers other than BPA-PC. We performed Car-Parrinello MD calcuations to predict the interaction energies of each of these molecules with a finite slab of nickel comprised of four atomic layers. Details of these results are presented in the contribution by L. Delle Site in these proceedings. Here, we summarize the results. We observed that at all relevant orientations, carbonic acid and propane are strongly repelled from nickel (111) when the distance between center of mass and first layer Ni atoms is below 3 Å. Both benzene and phenol strongly adsorb in a horizontal orientation at a distance of 2 Å from the top layer, with a strength of about 1 eV. This interaction is short-ranged, dying off above 3 Å, and strongly orientationally dependent, decaying nearly to zero if the angle between the normal of the surface of the ring and the normal of the nickel surface plane deviate in direction by more than about 45°.

The crucial step in incorporating this information into the 4:1 coarse-grained model is to realize two things. First, although internal phenylenes would apparently therefore like to stick to the nickel, they are sterically hindered by the strong repulsion felt by the adjacent isopropylidene. Second, the phenoxies at the ends of the chains are not sterically hindered because the carbonate can orient with one bridging oxygen at the surface and the others away, allowing the terminal phenoxy group to strongly bind without steric hindrance. However, the orientation of the ring is a sensitive determinant of the interaction strength, and this degree of freedom does not exist in the bulk 4:1 model.

2.3 The Dual Resolution BPA-PC Model

The solution to this problem is the dual-resolution model, depicted in Fig. 2. The essential feature of the dual-resolution model is that the terminal *carbonates* are resolved atomically. That is, instead of using a single c-type bead to represent the outermost carbonates, we use explictly one carbon atom, denoted "C," two bridging oxygen atoms, denoted "O_b, and a single carboxylic oxygen, denoted "O_C." The orientational degree of freedom we must access is determined by a vector joining the outer bridging oxygen of the terminal carbonate group and the center of the terminal phenyl ring. The direction of this vector is the "tilt" of the phenyl ring. Because we know the torsional barriers around this bond to be on the order of thermal energy at $T = 570$ K, we assume that rotation of the phenyl ring optimizes instantaneously relative to the tilt, and hence, we need only to resolve the carbonate group to access the tilt degree of freedom. Atomic-scale potentials taken from all-atom simulations of liquid diphenyl carbonate [Hahn99] are used to enforce all bonds, angles, and the one improper dihedral arising from this group.

The challenge in this dual-resolution scheme, as in any scheme that simultaneously resolves a single system with two or more levels of resolution, is handling the interface between the two levels of resolution. Here, this requires specifying how the inner O_b is bonded to its adjacent phenylene bead and how the outermost O_b is

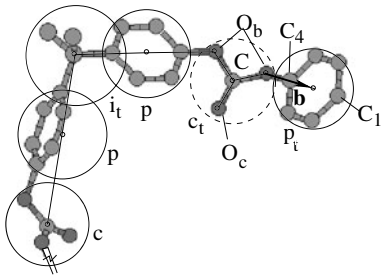


Fig. 2. Atomistic structure of a representative chain-end conformation of BPA-PC illustrating the dual-resolution scheme. The small black circles are mapping points in the scheme, and the large circles represent the excluded volume diameters of those points. The C_4-C_1 orientation vector, **b**, is shown. Lower-case letters denote coarse-grained bead designations: "p" is phenylene, "c_t" is the terminal carbonate, "p_t" is the terminal phenoxy, and "i_t" is the terminal isopropylidene. Capital letters denote atom type designations (see text). (See Plate 12 on Page 274)

bonded to its adjacent, chain-terminating phenyl, forming the phenoxy end. For this reason, the outermost phenylidene (p in Fig. 2 owns a mapping point tethered to its center of mass, unlike other phenylidenes. Then, the p–O_b bond is enforced at its natural length by a stiff harmonic bond. The same applies for the bond between the outermost O_b and the terminal phenyl p_t.

The p–O_b–C and p_t–O_b–C angles are treated using harmonic potentials as well, as if both the p-type and p_t beads were carbon atoms. The two O_b–C torsional potentials also use the p-type beads as carbon atoms. The angle defined by the outermost isopropylidene, i_t, the inner bridging oxygen, and the carbonate carbon is constrained at its geometrically determined value of 180° using a stiff harmonic potential. Because the outermost phenylene now owns a tethered mapping point, the angle at the outermost isopropylidene is determined by the sequence c–i_t–O_b obeys a different distribution than the normal c–i–c angle, which is computed ahead of time using the same type of Monte-Carlo averaging as in the previous case.

Finally, new intermolecular interactions must be accounted for relative to the previous 4:1 case when the carbonate groups are resolved atomically. However, as a first step, we have chosen the simplest route where the carbon atom is treated in an excluded volume sense as a c-type bead, and the oxygens experience no excluded volume.

The benefit of this new scheme is that the vector connecting the outermost O_b and the terminal phenyl bead determines the orientation of the terminal phenoxy *if* we make the assumption that a negligble atomic-scale torsional barrier on the O_b-C_4 exists (C_4 is the carbon atom of the phenyl that is bound to the bridging oxygen of the carbonate group). Indeed, this barrier is less than 1 kT at our temperature of interest, which is 570 K [Hahn99]. Thus, we can include the orientation of this vector as a degree of freedom upon which we can apply a potential energy, and this allows to accurately account for both the short range and strong orientational dependencies of the phenoxy-nickel interaction. Most importantly, because we are not forced to

resolve the entire molecule atomically, we know that no matter how quickly the the system responds to the presence of the surface, the chains more than one repeat unit distance from the surface are locally equilibrated, and global equilibration is feasible.

The simulation system is a slit pore, and particles are confined to a central region by the walls. Particle-wall interactions are purely repuslive for internal beads, and are (111)-site-specific and orientationally dependent for the terminal phenoxy beads. The MD simulations are preformed in reduced Lennard-Jones units (1 σ = 4.41 Å) using a time step of 0.005 τ and a Langevin-type thermostat [Kre90] with a friction of 0.5 τ^{-1}. Time is measured in units of $\tau \equiv \sigma\sqrt{m/k_{\mathrm{B}}T}$, where m is the unit mass. Because of the stochastic forces introduced by the thermostats, the masses of the explicit C and O atoms are artifically increased to 1.0 (which is the mass of all beads in the system) to reduce their bond oscillation frequencies. In this sense, using "heavy" carbon and oxygen in the dual resolution scheme is an acceptable alternative to limiting the integration to much smaller time-steps. Our assumption is that using such heavy atoms does not greatly affect the equilibrium structure of the chains or the liquid. The systems were initialized by first growing the chains as phantom random walks in a box with periodic boundaries in all three dimensions. Periodicity was then turned off in the z-dimension, and particle position z-components were unfolded. Walls are then placed at extremes in z and slowly, over 5000 MD steps, brought toward one another until the desired density is acheived.

The results presented in Ref. [Abr03b] were extracted from a simulation system of 240 chains, each chain having 10 chemical repeat units. Here, we compare a selection of these results to those from a system of 400 chains, each having 20 chemical repeat units.

3 Results and Discussion

Here we compare a selection of results for chains of 10 and 20 repeat units. The first result we consider is the evolution of the phenoxy-surface potential energy, E_s. Shown in Fig. 3(a) is the E_s vs. time for the two runs, normalized by the number of adsorbed ends at equilibrium, $\langle n_s \rangle$. In both cases, $E_s(t)/\langle n_s \rangle$ saturates to a value of roughly 12 kT at a time of about 10^5 τ. There is apparently not a significant effect on the surface equilibration from doubling the chain length from 10 to 20 repeat units. Furthermore, the slow component in the relaxation of E_s is due to new net adsorption events, as can be seen in Fig. 3(b). Here, the planar density of adsorbed ends ρ_s is plotted vs. time. $\rho_s(t)$ saturates to equilibrium values at a time of about 10^5 τ, and the equilibrium value of ρ_s for chains of length 10 repeat units is roughly 1.5 times larger than that for chains of length 20 repeat units. This results is somewhat surprising because the *volume* density of chain ends in the system with 20-repeat-unit chains is half that of the system with 10-repeat-unit chains. The longer chains can more easily adsorb ends than the shorter chains because tethering the end of a 20-repeat-unit chain sacrifices a smaller fraction of that chain's conformational entropy than would tethering the end of a 10-repeat-unit chain.

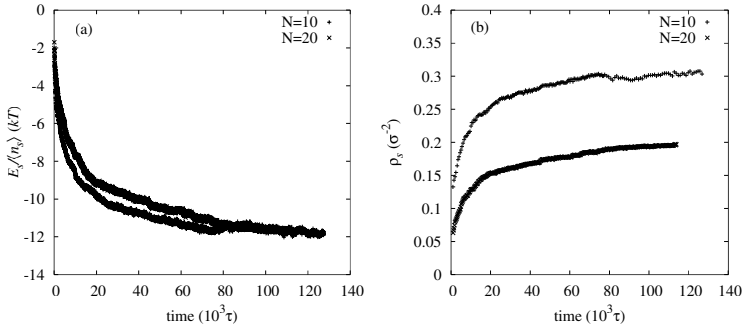


Fig. 3. (a) Phenoxy adsorption energy vs. time, normalized by the final number of adsorbed phenoxies, for systems of 10 and 20-repeat-unit chains. (b) Surface density of adsorbed phenoxies vs. time, again for systems of 10 and 20-repeat-unit chains.

Next, we consider the equilibrium structure and orientation of the chains. As presented in Ref. [Abr03b], the chain-end sticking for 10-repeat-unit BPA-PC resulted in a unique two-layered structure in chain packing and orientation as a function of distance from the wall. The dominant length scale in this structure is the unperturbed root-mean-square radius of gyration, R_g, which was about 20 Å for the 10-repeat-unit chains. It was observed that the chains with centers of mass less than about $2R_g$ from the surface were divided clearly into two populations: those with centers of mass between 0 and R_g from the surface (layer I) had both ends adsorbed, while those between R_g and $2R_g$ (layer II) had only one end each adsorbed and were strongly stretched along the direction normal to the wall. This is depicted schematically in Fig. 4. Moreover, the chains in layer I were flattened in the direction z normal to the wall, and chains in layer II were strongly stretched in z. The shapes of the chains are quantified in terms of the zz components of each chain's gyration tensor $\mathbf{S}$, where the trace of this tensor is R_g.

In the case of chains of 20 repeat units, the same general conclusions hold, as shown in Fig. 5. For these chains, R_g is about 29 Å. In Fig. 5(a), we plot both the density of chain centers-of-mass and S_{zz} as functions of distance from the wall, z. At a particular value of z, S_{zz} is the average of all chain S_{zz}'s for chains with centers of mass between z and $z + \delta z$. This data was averaged over the equilibrated configurations with a lateral resolution of δz=0.031 σ. Flattened chains in layer I and stretched chains in layer II are evident. In Fig. 5(b), we show the breakdown of the chains by the number of ends adsorbed. Here we see that most chains in layer I have both ends adsorbed, while most in layer II have only a single end adsorbed. The delineation between populations of chains with single- and double-ended adsorption is not as clear as for the shorter chains, but the same length scaling applies: Those chains with centers of mass less than one bulk R_g from the wall have both ends stuck and are somewhat flattened, while those between one and two R_g have a single end and are stretched. The degree of flattening and stretching is not as pronounced for the 20-repeat-unit chains as for the 10-repeat-unit chains, because each chain has

Fig. 4. Schematic representation of the molecular structure of a BPA-PC melt adjacent to a Ni wall, showing double-ended adsorbed, single-ended adsorbed, and non-adsorbed chains. (See Plate 13 on Page 275)

twice as many internal degrees of freedom upon which to call to relieve the stress associated with adsorbing a chain end to the surface while packing densly among other chains.

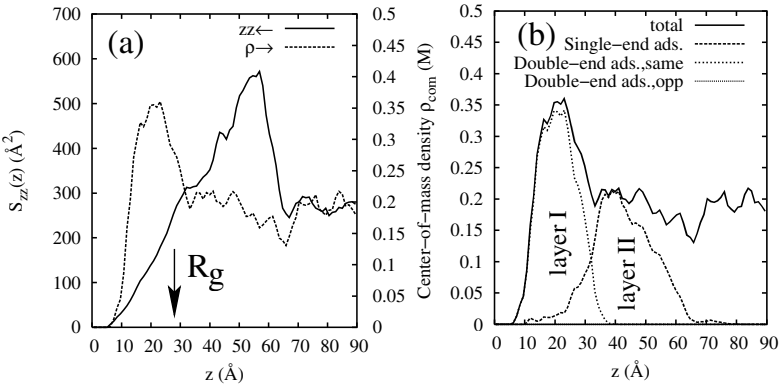


Fig. 5. Simulation results for 20-repeat-unit BPA-PC: (a) zz component of the lab-frame mean depth-resolved gyration tensor vs. distance from wall, and chain center of mass density vs. distance from wall. (b) Breakdown of chains into those with single- double-ended adsorption.

4 Summary, Conclusions, and Outlook

We have reported on the application of our dual-resolution model for BPA-PC-nickel interfaces to a system with longer BPA-PC molecules than in the original

paper [Abr03b]. We observe that the relaxation behavior of the surface interaction energy is not sensitive to chain length. Furthermore, equilibrium chain structure and orientation is qualitatively the same for the two chain lengths considered (10 and 20 chemical repeat units). For both cases, the liquid structure near the wall is divided into two layers: the outermost layer (immediately adjacent to the wall) contains flattened chains with both ends adsorbed, while the innermost layer (immediately adjacent to bulk liquid BPA-PC) contains stretched chains with single-ends adsorbed.

Our technique of using insight from *ab initio* to construct a minimalistic yet realistic coarse-grained model for a synthetic polymer interacting with a specific metal surface appears to be a promising technique which could be applied to many other specific systems. It should be noted here, however, that we have taken advantage of the unique interplay between sterics and energetics for BPA-PC-Ni, requiring that only chain ends be atomically resolved. In principle, one would rather have a technique which, when applied to any particular system with an interface or other interesting structure with atomic resolution in at least one dimension, would decide for itself what molecular fragments must be atomically resolved and what fragments may remain coarsened. This would, in effect, be a means to conduct embedded full-blown atomic MD simulations with realistic boundary conditions that optimize on-the-fly. Such ideas are not new in the area of simulations of crystalline materials, where one has the luxury of a reference crystal structure upon which a coarsened mesh can be constructed (for example, Ref. [Bro99]). It is also not new in the area of hybrid quantum/molecular mechanics models (QM/MM), in which a small region of a few atoms is treated quantum mechanically while being embedded in a larger system of atoms treated empirically (for example, Ref. [Eic99]). The idea, however, of "adaptive inhomogeneous coarse-graining" for molecular *liquids* expressed above is new, and introduces many unique considerations not encountered before in other multiresolution techniques. It is this area which we are devoting our most current research efforts.

5 Acknowledgments

The author is indebted to L. Delle Site and K. Kremer of the Max-Planck-Institute for Polymer Research. This work was supported by the BMBF under Grant No. 03 N 6015, and the Bayer Corp.

References

[Abr03a] ABRAMS, C. F., AND KREMER, K. Combined coarse-grained and atomistic simulation of liquid bisphenol-A-Polycarbonate: Liquid packing and intramolecular structure. *Macromolecules 36* (2003), 260–267.

[Abr03b] ABRAMS, C. F., SITE, L. D., AND KREMER, K. Dual-resolution coarse-grained simulation of the bisphenol-a-polycarbonate/nickel interface. *Phys. Rev. E 67* (2003), 021807.

[Bas00] BASCHNAGEL, J., BINDER, K., DORUKER, P., GUSEV, A. A., HAHN, O., KREMER, K., MATTICE, W. L., MÜLLER-PLATHE, F., MURAT, M., PAUL, W., SANTOS, S., SUTER, U. W., AND TRIES, V. Bridging the gap between atomistic and coarse-grained models of polymers: Status and perspectives. *Adv. Polym. Sci. 152* (2000), 41–156.

[Bro99] BROUGHTON, J. Q., ABRAMAM, F. F., BERNSTEIN, N., AND KAXIRAS, E. Concurrent coupling of length scales: Methodology and application. *Phys. Rev. B 60*, 4 (1999), 2391–2403.

[Eic99] EICHINGER, M., TAVAN, P., HUTTER, J., AND PARRINELLO, M. A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields. *J. Chem. Phys. 110*, 21 (1999), 10452–10467.

[Hahn01] HAHN, O., DELLESITE, L., AND KREMER, K. Simulation of polymer melts: From spherical to ellipsoidal beads. *Macromol. Th. Sim. 10*, 4 (2001), 288–303.

[Hahn99] HAHN, O., MOONEY, D. A., MÜLLER-PLATHE, F., AND KREMER, K. A new mechanism for penetrant diffusion in amorphous polymers: Molecular dynamics simulations of phenol diffusion in bisphenol-A-polycarbonate. *J. Chem. Phys. 111*, 13 (1999), 6061–6068.

[Kre90] KREMER, K., AND GREST, G. S. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *J. Chem. Phys. 92*, 8 (1990), 5057–5086.

[Kre01] KREMER, K., AND MÜLLER-PLATHE, F. Multiscale problems in polymer science: Simulation approaches. *MRS Bulletin 26*, 3 (2001), 205–210.

[Mor88] MORBITZER, L., AND GRIGO, U. Die wichtigsten polymerphysikalischen aspekte des polycarbonats. *Ange. Makromol. Chem. 162*, 2753 (1988), 87–107.

[M-P02] MÜLLER-PLATHE, F. Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back. *ChemPhysChem 3*, 9 (2002), 754–769.

[Paul91] PAUL, W., BINDER, K., KREMER, K., AND HEERMANN, D. W. Structure property correlation of polymers, a Monte-Carlo approach. *Macromolecules 24*, 23 (1991), 6332–6334.

[Del02] SITE, L. D., ABRAMS, C. F., ALAVI, A., AND KREMER, K. Polymers near metal surfaces: Selective adsorption and global conformations. *Phys. Rev. Lett. 89*, 15 (2002), 156103.

[Tsc98a] TSCHÖP, W., KREMER, K., BATOULIS, J., BÜRGER, T., AND HAHN, O. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polymer. 49* (1998), 61–74.

[Tsc98b] TSCHÖP, W., KREMER, K., HAHN, O., BATOULIS, J., AND BÜRGER, T. Simulation of polymer melts. II. From coarse-grained models back to atomistic description. *Acta Polymer. 49* (1998), 75–79.

[Wee71] WEEKS, J. D., CHANDLER, D., AND ANDERSEN, H. C. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys. 54*, 12 (1971), 5237–5247.

Stress and energy flow field near a rapidly propagating mode I crack

Markus J. Buehler[1], Farid F. Abraham[2], and Huajian Gao[1]

[1] Max Planck Institute for Metals Research, Heisenbergstr. 3, D-70569 Stuttgart, email m.buehler@mf.mpg.de (M. Buehler) hjgao@mf.mpg.de (H. Gao)
[2] IBM Almaden Research Center farid@almaden.ibm.com (F. Abraham)

Summary. Crack branching and instability phenomena are believed to be closely related to the circumferential or hoop stress in the vicinity of the crack tip. In this paper we show that the hoop stress around a mode I crack in a harmonic solid becomes bimodal at a critical speed of about 73 percent of the Rayleigh speed, in agreement with the continuum mechanics theory. Additionally, we compare the energy flow field predicted by continuum theory with the solution of molecular-dynamics simulations and show that the two approaches yield comparable results for the dynamic Poynting vector field. This study exemplifies joint atomistic and continuum modelling of nanoscale dynamic systems and yields insight into coupling of the atomistic scale with continuum mechanics concepts.

1 Introduction

Many phenomena associated with rapidly propagating cracks are not thoroughly understood. Experimental work [FGMS91] and computer simulations [ABRR94, ABRX97] have shown that initially straight cracks start to become unstable upon a critical speed of about 30 percent of their limiting speed, the Rayleigh velocity [Fre90]. In contrast, it was proposed based on the linear elastic continuum theory that cracks become unstable at about 75 percent of the Rayleigh speed (speed of surface waves) [Yof51, Fre90].

Up to date, it remains unclear what is the governing stress measure governing branching and crack instabilities at the atomic scale. Continuum formulations often use a material instability criterion similar to the principal stress to determine in which direction the material will fracture (e.g. [KG98]). It remains controversial if such a criterion can be accepted at the atomic scale, as the breaking of single atomic bonds is expected to govern the dynamics of cracks. Yoffe [Yof51] proposed that crack propagation should become unstable when the circumferential or hoop stress near a crack tip has a maximum at angles off the propagation direction. She showed that for low velocities, the hoop stress has a maximum ahead of the crack tip, but for velocities larger than about 73 percent of the Rayleigh speed, the hoop stress maximum is at off-angles of about 60 degrees, potentially causing the crack to change direction.

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

Atomistic simulations can successfully address many issues of dynamic fracture by providing an *ab initio* description of the fracture process [AWG+02a]. Here we use large-scale atomistic simulations to study the near-crack elastic fields in mode I dynamic fracture and to compare the discrete atomistic and the continuum mechanics viewpoints. The studies reported here are the first in a series of computer experiments where the crack instability is investigated. The main objective of this work is to show that the continuum mechanics prediction of a bimodal hoop stress field is reproduced quantitatively in molecular dynamics (MD) simulations. Since we focus on the deformation field near rapidly propagating cracks even beyond the instability speed, we constrain the crack to propagate along a one-dimensional prescribed fracture path modeled by a weak Lennard-Jones cohesive bonding. In the rest of the slab, atomic bonds are described by a harmonic potential and never break.

In the continuum theory, the stress field in the vicinity of the crack tip is given by the asymptotic solution [Yof51, Bak62, Fre90]. With $K_I(v)$ as the dynamic stress intensity factor,

$$\sigma_{ij}(v,\Theta) = \frac{K_I(v)}{\sqrt{2\pi r}}\,\Sigma_{ij}(v,\Theta) + \sigma_{ij}^{(1)} + O(1). \tag{1}$$

The functions $\Sigma_{ij}(v,\Theta)$ represent the variation of stress components with angle Θ for any value of crack speed v [Fre90]. The values of $\sigma_{ij}^{(1)}$ and the first order contribution $O(1)$ are determined from the boundary conditions, and neglected in the remainder of this work since the first term dominates very close to the crack tip. Similar expressions exist for the particle velocities $\dot{u}_i$ [Bak62, Fre90]. The asymptotic field strongly depends on the crack velocity, and has universal character because it is independent of the details of applied loading.

The plan of this paper is as follows. We start with an elastic analysis of a Lennard-Jones lattice as well as a harmonic lattice. We then show that in MD simulations of cracks travelling in perfect harmonic lattices the hoop stress becomes bimodal at about 73 percent of Rayleigh speed, in agreement with the continuum theory. In addition, we report comparison of continuum theory with MD simulation of the strain energy field near the crack tip as well as the energy transport field near rapidly moving cracks.

2 Continuum and atomistic modeling of dynamic fracture

Recent two decades have witnessed increasing research on MD simulations in materials science. Using huge supercomputers to break tiny nano-crystals has become an unforeseen fruitful combination. (see, e.g. [dAY83, CY93, ABRR94, Mar99, MG95, AWG+02a, AWG+02b, YDPG02, RKL+02, BHG03, BAG03, BGH03b]). The success of atomistic simulations of brittle fracture is partly due to the fact that the time and length scales involved in brittle dynamic fracture is perfectly suitable for MD simulations. Fracture is a process occurring on the order of sound speeds and a crack moves through a crystal with nanometer dimensions in a few picoseconds.

The basis of our simulations is the atomic interactions of a simple rare-gas solid accurately described by the Lennard-Jones 12:6 (LJ) potential [AT89]

$$\phi(r) = 4\epsilon_0 \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]. \tag{2}$$

All quantities are expressed in reduced units. Lengths are scaled by the LJ-parameter σ, and energies are scaled by the parameter ϵ_0 which is the depth of the minimum of the LJ potential. The reduced temperature is $\mathrm{k}T/\epsilon_0$ with k as the Boltzmann constant. The mass of each particle is $m = 1$. Such a two-dimensional lattice behaves as a brittle solid [ABRR94, ABRX97]. A simulation result of a system with LJ atomic interactions is shown in Figure 1. After an initial phase where cleavage is mirror-like, the crack surface starts to roughen at about 30 percent of the Rayleigh speed. Eventually, the crack surface turns into a hackle region accompanied by emission of dislocations. The final speed of the crack is around 50 percent of the Rayleigh speed. These observations are in agreement with [ABRR94, ABRX97].

The LJ potential is highly nonlinear. To rule out nonlinear effects, we linearize the LJ potential around the equilibrium distance $r_0 = 2^{1/6} \approx 1.12246$ and define a harmonic potential

$$\phi(r) = a_0 + \frac{1}{2}k(r - r_0)^2 \tag{3}$$

with $k = 72/2^{1/3} \approx 57.15$ and $a_0 = -1$.

To avoid crack branching or wiggling as seen in Figure 1, we constrain the crack to propagate along a one-dimensional fracture path. The slab geometry and orientation of the two-dimensional hexagonal lattice is shown in Figure 2. The slab width is

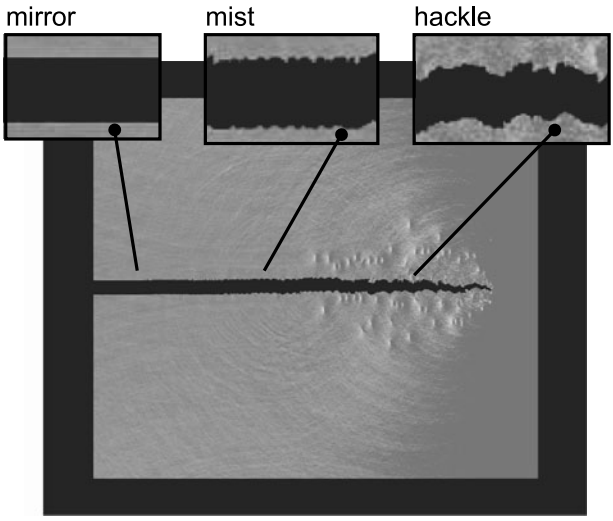


Fig. 1. Crack propagation in a LJ system as reported earlier in [ABRR94]. The plot shows the σ_{xx}-field and indicates the mirror-mist-hackle transition. (See Plate 14 on Page 275)

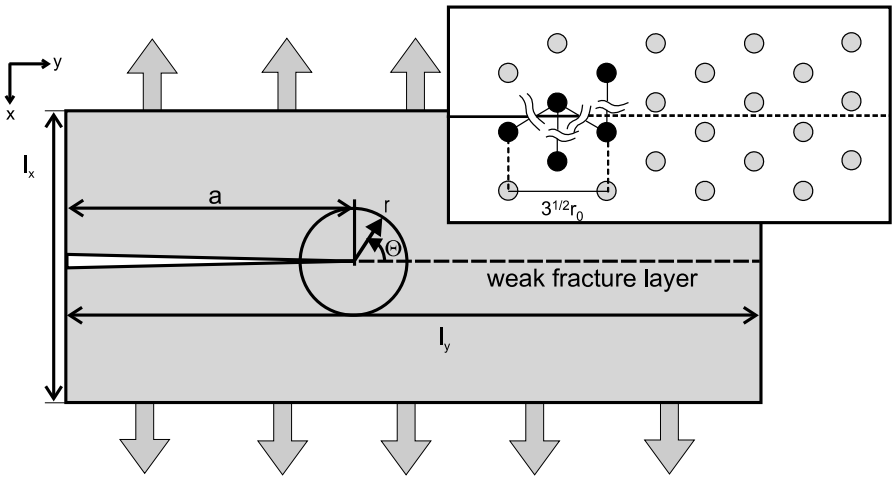


Fig. 2. Geometry of the simulation slab. The slab width is denoted by l_x, the slab length is l_y and the crack extension is a. The plot also depicts the crack orientation and the bond breaking process at the crack tip. (See Plate 15 on Page 276)

denoted by l_x, the slab length is $l_y \approx 3l_x$, the crack extension is a, and the crack is propagating in the y direction. To apply loading to the system, we displace the outermost rows of atoms in the slab according to a prescribed strain rate $\dot{\varepsilon} = 0.000,01$ and establish an initial linear velocity gradient in the slab. Note that the strain rate is given in reduced units defined by the LJ potential (see, e.g. [GHA01]) We use a higher loading rate of $\dot{\varepsilon} = 0.000,05$ in the studies of energy flow around the crack tip in order to drive the crack to higher velocities faster. Boundaries are held fixed at their current position to stop loading. The interactions across the weak layer are defined by the LJ potential given by equation (2), while the potential is defined by equation (3) in the rest of the slab. A horizontal slit of length a is cut midway along the left-hand vertical slab boundary, serving as a source for further crack extension. The crack is oriented orthogonal to the close-packed direction of the triangular lattice, and the slab is initialized with a temperature $T \approx 0$ K.

3 Elastic and fracture properties: From atomistics to continuum mechanics

The need to advance methods that combine atomistic and continuum analysis is becoming increasingly compelling with rapid advance in computational resources. In order to compare atomistic simulations with continuum analysis level, it is necessary to use methods that allow transition between the two levels of descriptions [KG98, ZKHG98, GK01]. Of particular interest is the relationship between interatomic potentials and associated elastic properties. Linking the two levels of descrip-

tion can be thought of as a homogenization procedure, since the lattice is a periodic rapidly varying function, and elastic properties are associated with a continuum.

A possible method of linking atomistic and continuum concepts is to use the Cauchy-Born rule [BH56, Hua02] which provides a relation between the energy created from a macroscopic strain field and the atomistic potential energy found in a stretched crystal lattice [Wei83, KG98, ZKHG98, GK01].

3.1 Elastic properties

For simple cases such as harmonic interactions in 2D, elastic properties can be calculated analytically. The linear spring potential given by equation (3) corresponds to the "ball-spring" model of solids and yields a plane-stress elastic sheet. Young's modulus E and shear modulus μ can be shown to be

$$E = \frac{2}{\sqrt{3}}k \approx 66, \quad \mu = \frac{\sqrt{3}}{4}k \approx 24.7. \tag{4}$$

It can further be shown that Poisson's ratio $\nu = 1/3$. Using the above values for elastic properties, the wave speeds can be straightforwardly obtained. The longitudinal wave speed can be calculated from the elastic properties as $c_l = \sqrt{3\mu/\rho}$, the shear wave speed is given by the square root of the ratio of the shear modulus μ to the density $\rho \approx 0.9165$, or $c_s = \sqrt{\mu/\rho}$. The speed of elastic surface waves, the Rayleigh speed, is given by $c_r \approx \beta c_s$, where $\beta \approx 0.9235$. The wave speeds are given by $c_l = 9$, $c_s \approx 5.2$ and $c_r \approx 4.82$.

Figure 3 shows numerical estimates of the elastic properties of a Lennard-Jones solid, where the elastic properties of the harmonic interactions are also included. The systems are loaded uniaxially in the two symmetry directions of the triangular lattice. The plot of the LJ system shows that the y direction requires a higher breaking strain than in the x direction (about 18 percent versus 12 percent). The tangent Young's modulus drops significantly from around 66 for small strain until it reaches zero when the solid fails. Poisson ratio remains around 1/3, but increases slightly when loaded in the x direction and decreases slightly when loaded in the orthogonal direction. The Young's modulus E with harmonic interactions increases with strain, and coincides with the value of the LJ solid for small strains. It can also be shown that the harmonic solid is isotropic for small strains, with Poisson's ratio close to 1/3 similar to the LJ solid.

3.2 Stress fields from atomistic simulations

In the continuum mechanics framework, stress and strain are two important quantities describing the state in an elastic body. A commonly used method to calculate stress from atomistic simulation results is the virial theorem [Tsa79], which is employed to define an atomic stress measure related, but not equivalent, to the Cauchy stress. In addition to stress, a frequently used quantity in continuum mechanics is strain, which describes the deformation of a body with respect to a

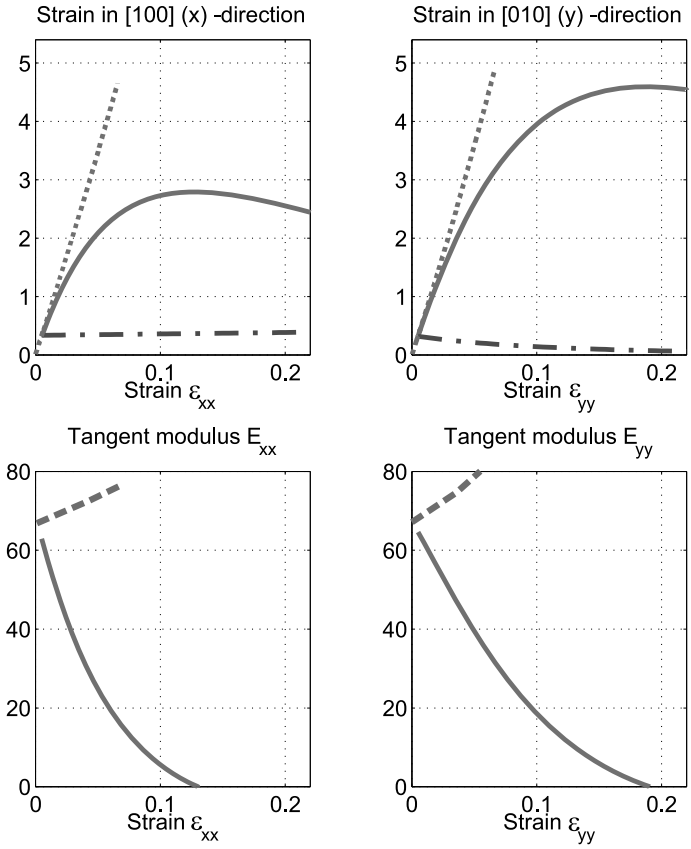


Fig. 3. The continuous lines show the elastic properties of the Lennard-Jones solid. The dashed lines show the elastic properties associated with the harmonic potential. The dash-dotted lines in the upper plots show Poisson's ratio.

reference configuration. This quantity can also be defined at the atomistic scale [Zim99, BGH03b, BAG03].

The stress field is a quantity containing information about the forces on the atoms in the material. The stress is obtained by considering variations of the energy with respect to application of a small strain. The virial stress is given by

$$\sigma_{ij} = \frac{1}{2} \sum_{\alpha,\beta} \left(-\frac{1}{r} \frac{\partial \phi}{\partial r} r_i r_j \mid_{r=r_{\alpha\beta}} \right). \tag{5}$$

We only consider the force part, excluding the part containing the effect of the velocity of atoms (the kinetic part), following a suggestion in [ZM02]. The hoop stress σ_Θ is determined by a transformation of the stress tensor into polar coordinates with origin at the current crack tip position. The principal stress is obtained by the usual

procedures [BC00]. The virial stress is strictly valid only under equilibrium conditions, and average over space and time is often required.

4 Simulation results

4.1 Angular variation of stress near the crack tip

We measure the angular variation of stress while the crack accelerates from low velocities to speeds close to the Ralyeigh speed. It is important to note that acceleration effects can severely change the resulting stress fields. Although the crack tip is regarded as inertia-less since it responds immediately to a change in loading or fracture surface energy, it takes time until the fields corresponding to a specific crack speed spread out. In fact, the fields spread out with the Rayleigh velocity behind, and with the shear wave velocity ahead of the crack. In other regions around the crack tip, the fields are reached in the long-time limit ($t \to \infty$) [BV81, Fre90, BGH03a].

We calculate the stress according to equation (5) for different crack speeds ranging from 0 to 87 percent of the Rayleigh speed. Figures 4 to 6 show the angular variation of σ_{xx}, σ_{yy} as well as σ_{xy}. Figure 7 shows the angular variation of the hoop stress σ_{Θ}. Figure 8 shows the angular variation of the maximum principal stress σ_1 near the crack tip. The snapshots are taken around a radius $R_0 \approx 11$. The continuous line is the corresponding continuum mechanics solution [Fre90]. It can be observed from the plots that the hoop stress becomes bimodal at a velocity of about 73 percent.

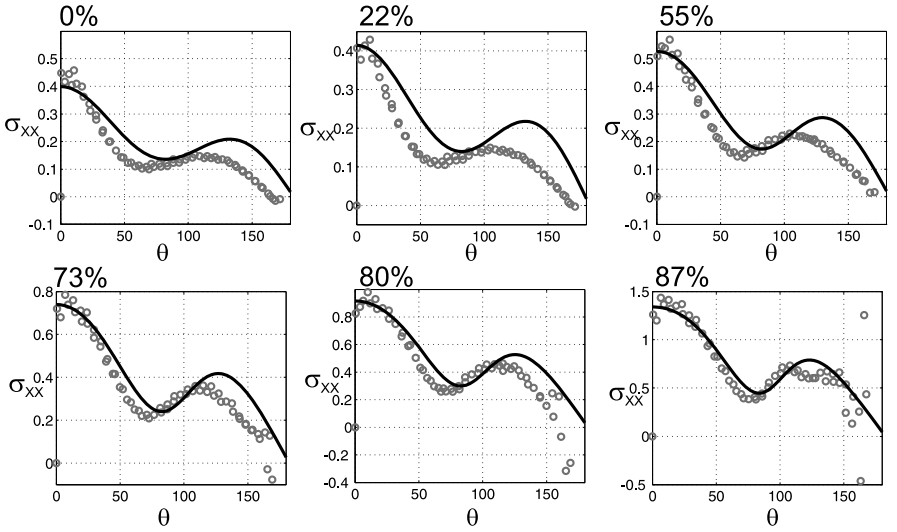


Fig. 4. Comparison between σ_{xx} from MD simulation with harmonic potential and the prediction of the continuum mechanics theory. The percentage refers to the ratio of crack speed with respect to the Rayleigh velocity.

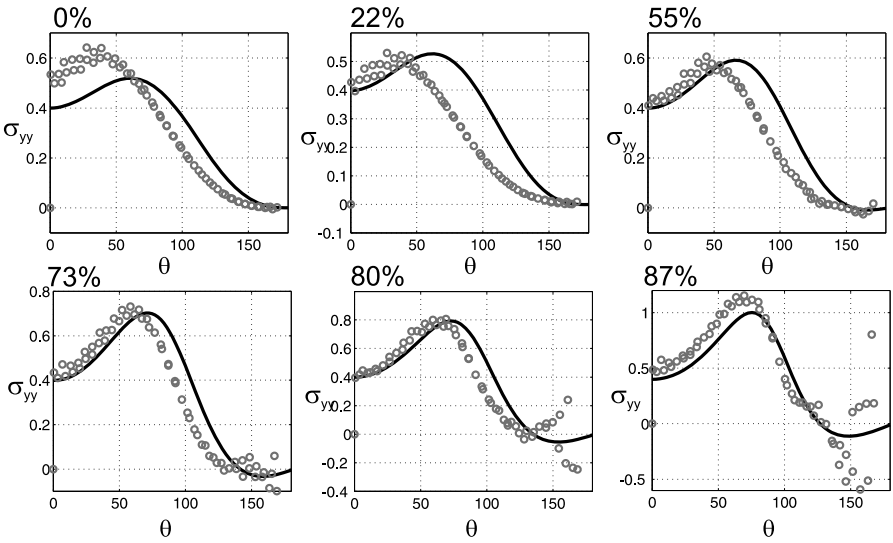


Fig. 5. Comparison between σ_{yy} from MD simulation with harmonic potential and the prediction of the continuum mechanics theory. The percentage refers to the ratio of crack speed with respect to the Rayleigh velocity.

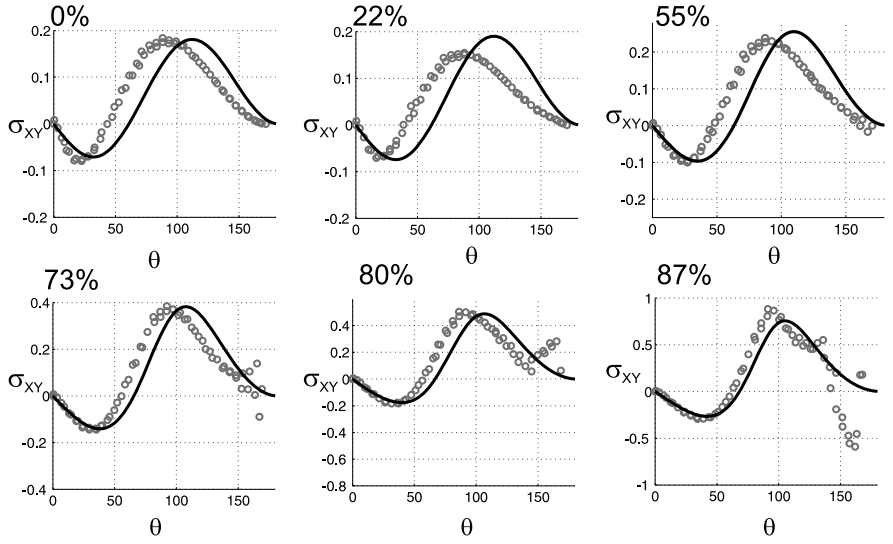


Fig. 6. Comparison between σ_{xy} from MD simulation with harmonic potential and the prediction of the continuum mechanics theory. The percentage refers to the ratio of crack speed with respect to the Rayleigh velocity.

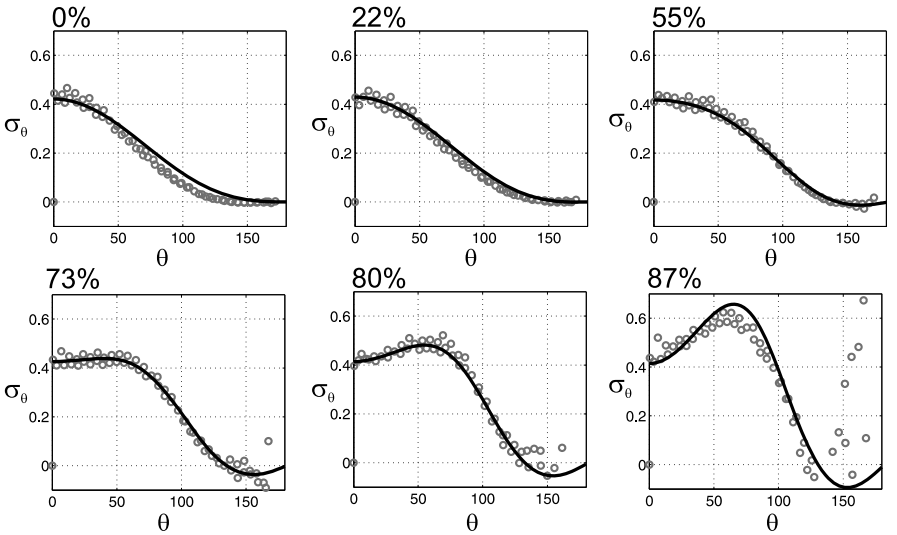


Fig. 7. Comparison between hoop stress from MD simulation with harmonic potential and the prediction of the continuum mechanics theory. The percentage refers to the ratio of crack speed with respect to the Rayleigh velocity.

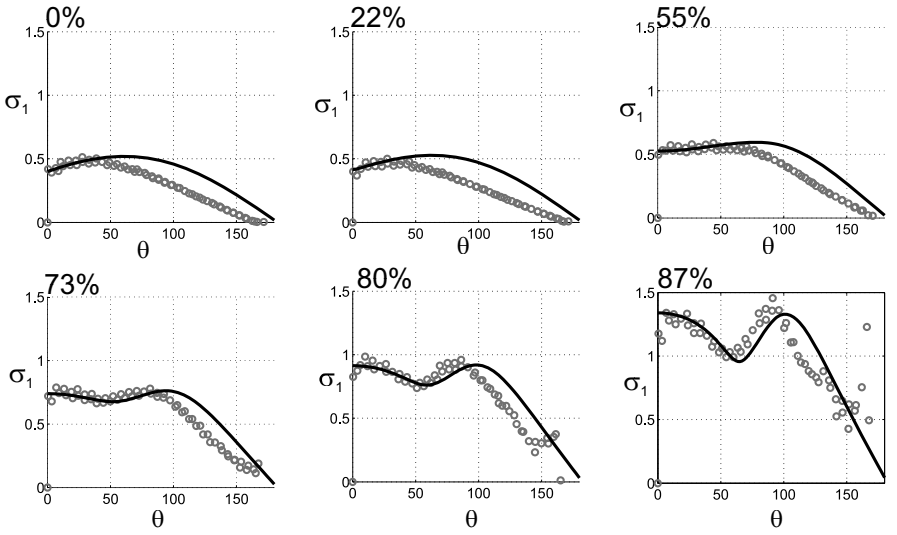


Fig. 8. Comparison between the maximum principal stress σ_1 from MD simulation with harmonic potential and the prediction of the continuum mechanics theory. The percentage refers to the ratio of crack speed with respect to the Rayleigh velocity.

4.2 Energy flow near a crack tip

Here we discuss the energy flow near a crack tip in MD simulations compared with the continuum theory [Fre90]. A similar study has been reported recently [FPG+02]. In contrast to our treatment of the dynamic Poynting vector for steady-state cracks at high velocities (in analogy to the discussion in [Fre90]), the authors in [FPG+02] only consider the static Poynting vector to study energy radiation of rapidly moving cracks.

The dynamic Poynting vector for a crack moving at velocity v can be expressed as

$$P_i = \sigma_{ij}\, \mathbf{n_j}\, \dot{u}_i + (U+T) v\, \mathbf{n_1}, \tag{6}$$

where $\mathbf{n_i}$ the basis vector in the i-th direction, $\mathbf{n_1}$ denoting the crack propagation direction. The kinetic energy is given by $T = \frac{1}{2}\rho \dot{u}_i \dot{u}_i$, and the strain energy density for an isotropic medium is given by [BC00]

$$U = \frac{1}{2E}\left(\sigma_{11}^2 + \sigma_{22}^2 - 2\nu\sigma_{11}\sigma_{22} + 2(1+\nu)\sigma_{12}^2\right). \tag{7}$$

The magnitude of the dynamic Poynting vector is calculated as $P = \sqrt{P_1^2 + P_2^2}$, and can be identified as a measure for the local energy flow [BAG03].

Figure 9 (a) shows the strain energy field near the crack tip predicted by both the continuum theory prediction and the MD simulation result. Figure 9 (b) shows the magnitude of the dynamic Poynting vector field. Figure 10 shows in panel (a) the continuum mechanics prediction, and in panel (b) the MD simulation result of the

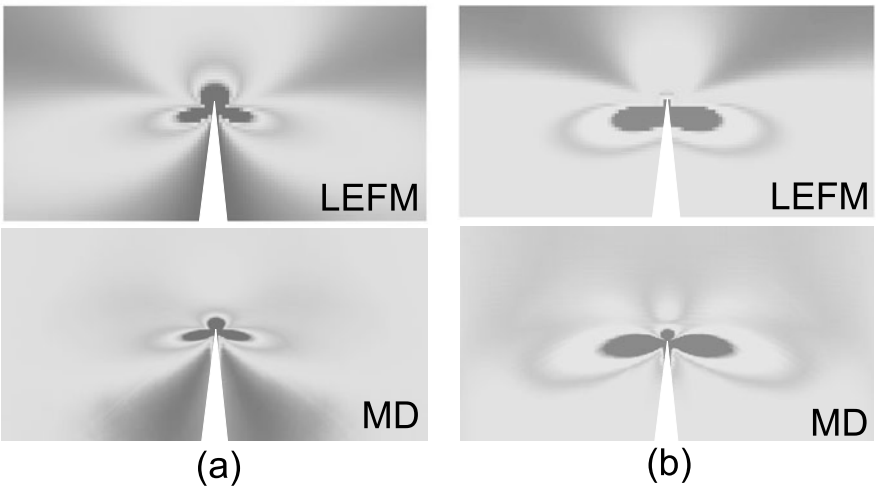


Fig. 9. Potential energy field and magnitude of the dynamic Poynting vector. (a) Potential energy field near a crack close to the Rayleigh speed. (b) Energy flow near a rapidly propagating crack. This plot shows the magnitude of the dynamic Poynting vector in the vicinity of a crack propagating at a velocity close to the Rayleigh speed. (See Plate 16 on Page 276)

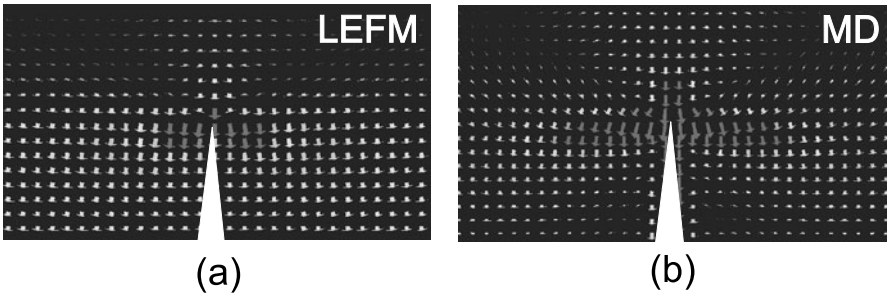


Fig. 10. Energy flow near a rapidly propagating crack. This plot shows (a) the continuum mechanics prediction, and (b) the MD simulation result of the dynamic Poynting vector field in the vicinity of the crack tip, for a crack propagating close to the Rayleigh speed. (See Plate 17 on Page 277)

dynamic Poynting vector field in the vicinity of the crack tip, for a crack propagating close to the Rayleigh speed.

5 Discussion and outlook

The main result of this paper is that the hoop stress near a crack tip is reproduced well in the atomistic simulations. The Yoffe prediction that the hoop stress field becomes bimodal above 73 percent of the Rayleigh speed is found in MD simulation results as well [Yof51, Fre90]. This could be verified in Figure 7. Further, the results shown in this paper suggest that the agreement of MD simulations and continuum mechanics is generally good, as it is shown for the stress tensor components σ_{xx}, σ_{yy} and σ_{xy}. We observe disagreement at large angles $\Theta > 150$ degrees, probably due to surface effects in the atomistic simulations.

In Figure 5 we observe that for σ_{xx}, the shape of $\sigma_{xx}(\Theta)$ is qualitatively reproduced well over the entire velocity regime between 0 and 87 percent of Rayleigh speed. However, the angles of the local maxima and minima are shifted to slightly smaller values compared to the theory prediction. Figure 5 illustrates that the shift of the maximum in the $\sigma_{yy}(\Theta)$ curve from about 60 degrees to about 80 degrees is reproduced only qualitatively. For low velocities the maximum is found at lower angles around 40 degrees, but it approaches the value of the continuum theory at higher velocities. At 87 percent of Rayleigh speed, the difference is only a few degrees. The shear stress σ_{xy} shown in Figure 6 also agrees qualitatively with the continuum theory. As in the previous cases, the angles of local minima and maxima are shifted to lower values in the simulation, but the agreement gets better when the crack velocity is faster. Even though we see small deviations in σ_{xx}, σ_{yy} and σ_{xy}, the hoop stress σ_{Θ} agrees quantitatively with the continuum theory as shown in Figure 7. The angles of the maxima and minima during crack acceleration agree well with theory. However, the angles of the maxima and minima of the maximum principle stress shown in Figures 8 are also shifted to slightly lower values. However, we observe that two local maxima and one local minima develop at a velocity of about 73 percent in quan-

titative agreement with continuum theory ("trimodal structure"). The magnitude of the local maxima and minima also agree quantitatively.

Our results are in line with a previous study where a snapping bond weak layer was used and a systematic comparison of the close-crack tip deformation field with continuum mechanics was reported [BGH03b]. The weak LJ layer approach could be considered advantageous for modeling crack tip field since crack acceleration can be much smoother. Using a snapping bond weak layer, the minimum crack speed is about 50 percent of Rayleigh speed with discrete jumps in the propagation speed as the crack accelerates [BGH03b].

We find that the crack acceleration can cause significant changes in the near tip fields. The slower the crack acceleration, the better agreement we find between the measured and predicted fields. This can be explained by the fact that it takes time to establish the asymptotic field at a fixed position near a crack tip [BV81, Fre90, BGH03a].

The analysis of the potential energy field near a crack close to the Rayleigh speed agrees qualitatively with the prediction by the continuum mechanics theory. As Figure 9 (a) shows, in both theory and computation the field clearly shows two three local maxima with respect to the angular variation ("trimodal structure"), similar to the principal stress field. At larger distances away from the crack tip we observe that other stress terms begin to dominate in the simulation, so the distribution of the potential energy deviates from the prediction by theory. As is expected since only the first term of equation (1) is considered, these contributions are missing in the continuum solution. Similar observations also hold for the magnitude of the dynamic Poynting vector, as it can be verified in Figure 9 (b). The dynamic Poynting vector field calculated by MD is also in reasonable agreement with the continuum mechanics prediction. This could be verified in Figure 10. In both theory (Figure 10 (a)) and molecular dynamics calculation (Figure 10 (b)), the orientation of the dynamic Poynting vector is dominated by the direction opposite to crack motion. The vector field seems to bow out around the crack tip, an effect that is more pronounced in the simulation than predicted by theory. Also, the flow ahead of the crack is larger in simulation than predicted by theory. At the free surface of the crack, the measurement from the simulation and the prediction by theory show differences. This could be based on the fact that the continuum theory does not treat surface effects properly, in particular short-wave length Rayleigh waves (see also discussion in [FPG$^+$02]).

Finally, we note that the virial expression of the stress tensor is classically thought to be only valid under equilibrium conditions [Tsa79]. Our results in Figures 4-8 show that it is approximately valid even under dynamic fracture conditions which are far from equilibrium.

Work in the immediate future will involve crack dynamics in unconstrained harmonic solids in order to study whether the hoop stress is the governing stress for crack branching and instability. Further research will be directed in investigating the role of hyperelasticity in crack dynamics [BAG03].

Acknowledgements

The simulations were carried out in the Max Planck Society Supercomputer Center in Munich. MJB acknowledges the generous support of ETH's CoLab during the *Workshop for Multiscale Modeling and Simulation*. FFA acknowledges the generous support provided by the Humboldt Research Award for Senior U.S. Scientists.

References

[ABRR94] F.F. Abraham, D. Brodbeck, R.A. Rafey, and W.E. Rudge. Instability dynamics of fracture: A computer simulation investigation. *Phys. Rev. Lett.*, 73(2):272–275, 1994.

[ABRX97] F.F. Abraham, D. Brodbeck, W.E. Rudge, and X. Xu. A molecular dynamics investigation of rapid fracture mechanics. *J. Mech. Phys. Solids*, 45(9):1595–1619, 1997.

[AT89] M.P. Allen and D.J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 1989.

[AWG+02a] F.F. Abraham, R. Walkup, H. Gao, M. Duchaineau, T.D. de la Rubia, and M. Seager. Simulating materials failure by using up to one billion atoms and the world's fastest computer: Brittle fracture. *PNAS*, 99(9):5788–5792, 2002.

[AWG+02b] F.F. Abraham, R. Walkup, H. Gao, M. Duchaineau, T.D. de la Rubia, and M. Seager. Simulating materials failure by using up to one billion atoms and the world's fastest computer: Work-hardening. *PNAS*, 99(9):5783–5787, 2002.

[BAG03] M.J. Buehler, F.F. Abraham, and H. Gao. Hyperelasticity governs dynamic fracture at a critical length scale. *Nature*, 426:141–146, 2003.

[Bak62] B.R. Baker. Dynamic stresses created by a moving crack. *Journal of Applied Mechanics*, 29:567–578, 1962.

[BC00] A. Boresi and K.P. Chong. *Elasticity in Engineering Mechanics*. Wiley-Interscience, New York, 2nd edition, 2000.

[BGH03a] M.J. Buehler, H. Gao, and Y. Huang. Continuum and atomistic studies of a suddenly stopping supersonic crack. *Computational Materials Science*, 28(3-4):385–408, 2003.

[BGH03b] M.J. Buehler, H. Gao, and Y. Huang. Continuum and atomistic studies of the near-crack field of a rapidly propagating crack in a harmonic lattice. *Theor. Appl. Fract. Mech.*, in press, 2003.

[BH56] M. Born and K. Huang. *Dynamical Theories of Crystal Lattices*. Clarendon, Oxford, 1956.

[BHG03] M.J. Buehler, A. Hartmeier, and H. Gao. Atomistic and continuum studies of crack-like diffusion wedges and dislocations in submicron thin films. *J. Mech. Phys. Solids*, 51:2105–2125, 2003.

[BV81] V.K. Kinra B.Q. Vu. Britle fracture of plates in tension – static field radiated by a suddenly stopping crack. *Engrg. Fracture Mechanics*, 15(1-2):107–114, 1981.

[CY93] K.S. Cheung and S. Yip. A molecular-dynamics simulation of crack tip extension: the brittle-to-ductile transition. *Modelling Simul. Mater. Eng.*, 2:865–892, 1993.

[dAY83] B. deCelis, A.S. Argon, and S. Yip. Molecular-dynamics simulation of crack tip processes in alpha-iron and copper. *J. Appl. Phys.*, 54(9):4864–4878, 1983.

[FGMS91] J. Fineberg, S.P. Gross, M. Marder, and H.L. Swinney. Instability in dynamic fracture. *Phys. Rev. Lett.*, 67:141–144, 1991.

[FPG+02] S. Fratini, O. Pla, P. Gonzalez, F. Guinea, and E. Louis. Energy radiation of moving cracks. *Phys. Rev. B*, 66(10):104104, 2002.

[Fre90] L.B. Freund. *Dynamic Fracture Mechanics*. Cambridge University Press, 1990.

[GHA01] H. Gao, Y. Huang, and F.F. Abraham. Continuum and atomistic studies of intersonic crack propagation. *J. Mech. Phys. Solids*, 49:2113–2132, 2001.

[GK01] H. Gao and P. Klein. Numerical simulation of crack growth in an isotropic solid with randomized internal cohesive bonds. *J. Mech. Phys. Solids*, 46(2):187–218, 2001.

[Hua02] K. Huang. On the atomic theory of elasticity. *Proc. R. Soc. London*, 203:178–194, 2002.

[KG98] P. Klein and H. Gao. Crack nucleation and growth as strain localization in a virtual-bond continuum. *Engineering Fracture Mechanics*, 61:21–48, 1998.

[Mar99] M. Marder. Molecular dynamics of cracks. *Computing in Science and Engineering*, 1(5):48–55, 1999.

[MG95] M. Marder and S. Gross. Origin of crack tip instabilities. *J. Mech. Phys. Solids*, 43(1):1–48, 1995.

[RKL+02] C.L. Rountree, R.K. Kalia, E. Lidorikis, A. Nakano, L. van Brutzel, and P. Vashishta. Atomistic aspects of crack propagation in brittle materials: Multimillion atom molecular dynamics simulations. *Annual Rev. of Materials Research*, 32:377–400, 2002.

[Tsa79] D.H. Tsai. Virial theorem and stress calculation in molecular-dynamics. *J. of Chemical Physics*, 70(3):1375–1382, 1979.

[Wei83] J.J. Weiner. Hellmann-feynmann theorem, elastic moduli, and the cauchy relation. *Phys. Rev. B*, 24:845–848, 1983.

[YDPG02] V. Yamakov, D. Wolf D, S.R. Phillpot, and H. Gleiter. Grain-boundary diffusion creep in nanocrystalline palladium by molecular-dynamics simulation. *Acta mater.*, 50:61–73, 2002.

[Yof51] E.H. Yoffe. The moving griffith crack. *Philosophical Magazine*, 42:739–750, 1951.

[Zim99] J. Zimmermann. *Continuum and atomistic modelling of dislocation nucleation at crystal surface ledges*. PhD thesis, Stanford University, 1999.

[ZKHG98] P. Zhang, P. Klein, Y. Huang, and H. Gao. Numerical simulation of cohesive fracture by the virtual-internal-bond model. *CMES-Computer Modeling in Engineering and Sciences*, 3(2):263–277, 1998.

[ZM02] M. Zhou and D.L. McDowell. Equivalent continuum for dynamically deforming atomistic particle systems. *Phil. Mag. A*, 82(13):2547–2574, 2002.

A Peierls Criterion for Deformation Twinning at a Mode II Crack

E. B. Tadmor

Department of Mechanical Engineering, Technion – Israel Institute of Technology, 32000 Haifa, Israel.

Summary. The theoretical criterion for deformation twinning (DT) at a crack tip due to Tadmor and Hai [TH03] is reviewed. The criterion quantifies the competition between slip and DT at a crack tip in terms of the critical stress intensity factors necessary to nucleate a dislocation and a two-layer microtwin there. The analysis is based on Rice's Peierls framework for dislocation emission from a crack tip [Ric92]. It is found that DT at a crack tip is controlled by a new material parameter named the "unstable twinning energy", which plays an analogous role for twinning as Rice's unstable stacking energy plays for slip. The derivation presented in this paper is for the simplest possible special case: a crack with a co-planar slip plane loaded in pure mode II shear along a crystallographic slip direction. The analysis is limited to a face-centered cubic (fcc) crystal. While this derivation contains all of the features in the full derivation in [TH03], the simpler geometry helps to clarify the underlying concepts. Some of the expressions derived for this special case are new.

1 Introduction

In this paper, the Peierls criterion for DT at a crack tip due to Tadmor and Hai [TH03] is reviewed. The criterion is based on Rice's Peierls framework for dislocation nucleation from a crack tip [Ric92].

Consider a crack with a slip plane intersecting the crack front (Fig. 1). The objective of the analysis is to obtain the critical stress intensity factors (SIFs) required to nucleate a dissociated dislocation and a deformation twin from the crack tip. A dissociated dislocation is formed by the emission of a leading partial dislocation followed by a trailing partial dislocation on the same plane. A twin nucleus is formed by the emission of a leading partial dislocation followed by another leading partial dislocation on the plane above the first partial. The competition between slip and twinning can then be formulated in terms of the competition between the two possible secondary partials after the leading partial has been emitted. This is demonstrated in Fig. 2.

This paper presents the derivation of the DT criterion for the simplest possible special case: the slip plane is coplanar with the crack plane (i.e. $\theta = 0$), the crack is

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

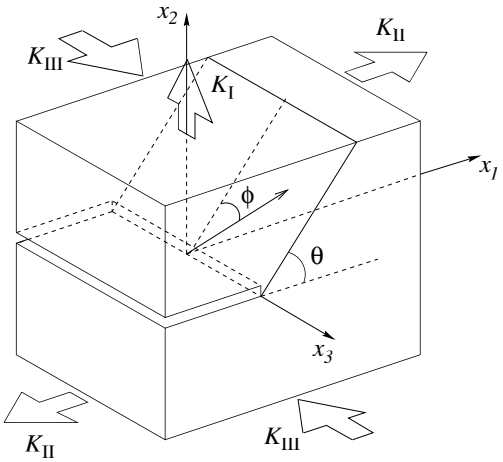


Fig. 1. A schematic diagram of the crack tip region showing the inclined slip plane and slip direction posited by the dislocation nucleation model. Reproduced from [TH03].

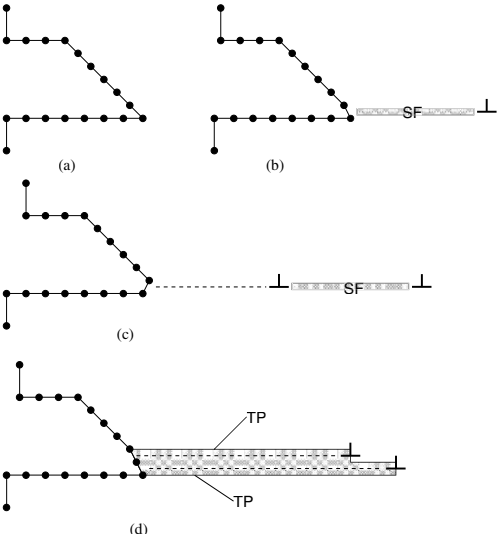


Fig. 2. The competition between slip and twinning at a crack tip. The atomically-sharp crack in (a) is loaded and emits a partial dislocation with a stacking fault (SF) in its wake in (b). As the load continues to increase, the crack will either emit the trailing partial and form a dissociated dislocation in (c), or emit a second leading partial on an adjacent plane forming a two-layer microtwin delimited by twin planes (TP) in (d). Reproduced from [TH03].

loaded in pure mode II, and the slip direction of the leading partial is oriented along the positive x_1 axis (i.e. $\phi_A = 0$). In an fcc crystal this implies that the x_1 lies along a $\langle 112 \rangle$ direction and the x_2 axis lies along a $\langle 111 \rangle$ direction. The trailing partial will form an angle of $\phi_B = 60^\circ$ with the x_1-axis. The objective of this derivation

is to clarify the main features of the theory (which are all present here) by reducing as much as possible complications associated with crystallography and loading. The derivation contains new expressions, but it is based entirely on Rice's derivation for the leading and trailing partials in [Ric92], and Tadmor and Hai's derivation for the twinning partial in [TH03].

2 Emission of the leading partial

The solution for the emission of the leading partial is obtained through application of the J-Integral [Ric68]. A displacement discontinuity $\delta = \delta_1 = u_1^+ - u_1^-$ is assumed to exist along the x_1-axis ahead of the crack tip (Fig. 3). Applying the Peierls concept, the shear stress along this line is taken to be a function of the local slip discontinuity, $\tau(x_1) = f(\delta(x_1))$. An interplanar potential $\Phi(\delta) = \int \tau d\delta$ may be defined, such that $\tau = d\Phi/d\delta$. Away from the slip plane the material is assumed to be linear elastic and isotropic with shear modulus μ and Poisson's ratio ν.

Rice [Ric92] pointed out that the potential $\Phi(\delta)$ used in the Peierls model is related (but not equal) to the generalized stacking fault interplanar potential $\Psi(\Delta)$ [Vit68]. The difference is that $\Psi(\Delta)$ is the energy obtained by rigidly displacing one half of a crystal relative to the other, where Δ is the disregistry of atoms across the cut plane. For a lattice undergoing a uniform shear γ outside of the slip plane, the disregistry and slip discontinuity are related by $\Delta = \delta + \gamma h$, where h is the interplanar spacing normal to the slip plane. It may be shown that the potentials $\Phi(\delta)$ and $\Psi(\Delta)$ are related according to $\Phi = \Psi - h\tau^2/2\mu$. An important consequence of this relation is that the potentials $\Phi(\delta)$ and $\Psi(\Delta)$ have the same extrema. In particular, the parameter γ_{us}, a maximum on the $\Phi(\delta)$ curve which plays an important role in the nucleation criterion, may be obtained directly from the corresponding $\Psi(\Delta)$ function.

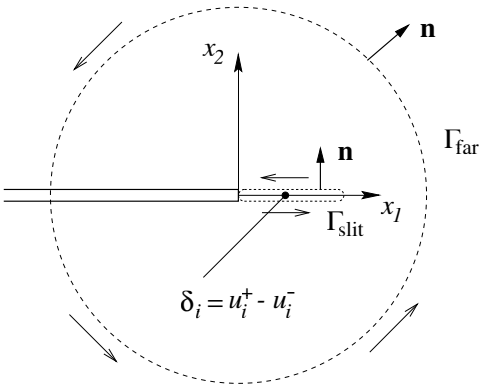


Fig. 3. A schematic diagram of the crack tip loaded in mode II with the slip direction coincident with the loading directions. The J-integral integration contours are indicated.

To solve the nucleation problem the path invariance property of the J-integral is utilized, namely that J, given by

$$J = \int_{\Gamma} [n_1 W - n_\alpha \sigma_{\alpha\beta} \partial u_\beta / \partial x_1] \, ds, \tag{1}$$

will be the same for any circuit Γ which does not traverse the crack or the slip zone ahead of it. In (1), $\mathbf{n}$ is the unit normal to the path Γ, W is the strain energy density, $\boldsymbol{\sigma}$ is the stress tensor, $\mathbf{u}$ is the displacement field and s is the arc length parameter. Two contours are defined: Γ_{far}, which is far removed from the crack tip where the effects of the nonlinearities associated with the slip can be neglected, and Γ_{slit}, which is coincident with the slipped zone ahead of the crack tip. By equating $J_{\text{far}} = J_{\text{slit}}$, a correspondence between the macroscopic field parameters and the microscopic properties of the slipped zone is obtained. Integrating (1) on the far contour yields the well-known result demonstrating the equivalence between J and the energy release rate G,

$$J_{\text{far}} = (1-\nu) K_{\text{II}}^2 / 2\mu \equiv G. \tag{2}$$

Integrating (1) on the slit contour turns out to also have a clear significance. The first term in (1) drops out, since $n_1 = 0$, with the result,

$$\begin{aligned} J_{\text{slit}} &= -\int_0^{\infty} \sigma_{21} \frac{\partial (u_1^+ - u_1^-)}{\partial x_1} dx_1 \\ &= -\int_0^{\infty} \tau[\delta(x_1)] \frac{\partial \delta}{\partial x_1} dx_1 = \int_0^{\delta_{\text{tip}}} \tau \, d\delta = \Phi(\delta_{\text{tip}}), \end{aligned} \tag{3}$$

where the relations $\tau = \sigma_{21}$ and $\delta = u_1^+ - u_1^-$ have been used, and where δ_{tip} is the slip discontinuity at the tip of the crack. Equating equations (2) and (3) gives,

$$(1-\nu) K_{\text{II}}^2 / 2\mu = \Phi(\delta_{\text{tip}}). \tag{4}$$

As K_{II} increases the energy stored in the slipped zone increases and the potential at the crack tip follows the $\Phi(\delta)$ curve. At the first maximum of $\Phi(\delta)$ the system loses stability and a dislocation is emitted. This maximum is defined as the unstable stacking energy γ_{us}. The critical SIF for the nucleation of the leading partial dislocation is thus,

$$K_{\text{II}}^1 = \sqrt{\frac{2\mu\gamma_{\text{us}}}{1-\nu}}, \tag{5}$$

where the superscript "1" indicates that this is the critical load for the emission of the first partial from the crack tip. As soon as $K_{\text{II}} = K_{\text{II}}^1$ the leading partial dislocation is nucleated and glides away from the crack tip leaving behind a stacking fault ribbon with energy per unit area γ_{sf}. This sequence of events is demonstrated in Fig. 4 where the interplanar potential $\Psi(\Delta)$ is plotted for slip along the leading partial direction. The figure also includes the atomic configurations associated with five points along the curve. The $\Psi(\Delta)$ curve was computed using an Embedded Atom Method (EAM)

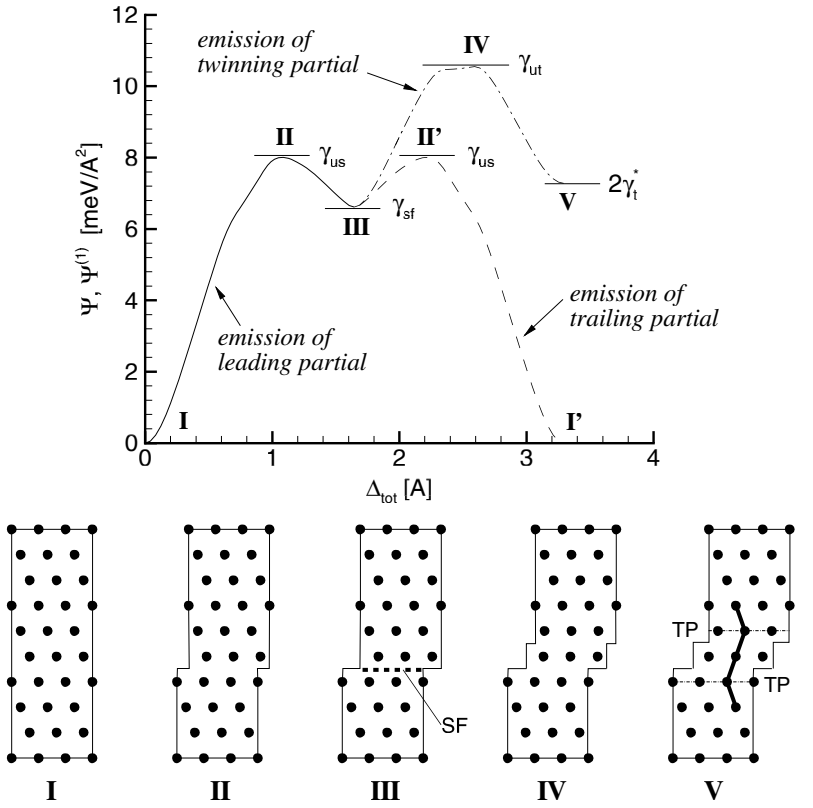


Fig. 4. The interplanar potential $\Psi(\Delta)$ for aluminum for the emission of a leading partial (solid line), trailing partial (dashed line) and a twinning partial (dash-dot line) along with the atomic arrangements associated with points I to V on the $\Psi(\Delta)$ curve. Reproduced from [TH03].

[DB83] potential for aluminum due to Ercolessi and Adams [EA93]. The first part of the curve from I to III is associated with the emission of the leading partial dislocation. The system will lose stability at point II where $\Phi(\delta_{\rm tip}) = \Psi(\Delta_{\rm tip}) = \gamma_{\rm us}$ and emit a partial dislocation. Following the emission, the slip plane behind the partial contains a stacking fault and remains at an elevated energy level $\gamma_{\rm sf}$ (point III).

The nucleated partial dislocation moves away from the crack tip and settles at an equilibrium distance r_A. This distance is determined by the balance of image and stacking fault forces drawing the dislocation back to the crack tip and the Peach-Koehler force driving the dislocation away,

$$\frac{K_{\rm II} b_A}{\sqrt{2\pi r_A}} = \gamma_{\rm sf} + \frac{\mu b_A^2}{4\pi(1-\nu) r_A}, \tag{6}$$

where b_A is the Burgers vector of the leading partial (in an fcc crystal $b_A = a_0/\sqrt{6}$ where a_0 is that lattice parameter). The equilibrium equation in (6) has two roots: a point of unstable equilibrium very close to the origin and a point of stable equilibrium normally at $r_A \gg b_A$. The larger root is the one of interest,

$$r_A = \frac{K_{\rm II}^2 b_A^2}{8\pi\gamma_{\rm sf}^2}\left[1+\sqrt{1-\frac{2\mu\gamma_{\rm sf}}{(1-\nu)K_{\rm II}^2}}\ \right]^2. \tag{7}$$

3 Emission of the trailing partial

As $K_{\rm II}$ continues to increase the incipient slip associated with the trailing partial may begin to form. The critical SIF for the emission of the trailing partial is obtained in a similar manner to the emission of the leading partial. There are, however, several important differences:

(i) The leading partial dislocation that was emitted earlier shields the crack tip and reduces the stress intensity factor from $K_{\rm II}$ to $K_{\rm II}^*$,

$$K_{\rm II}^* = K_{\rm II} - \frac{\mu b_A}{(1-\nu)\sqrt{2\pi r_A}} = \sqrt{K_{\rm II}^2 - \frac{2\mu\gamma_{\rm sf}}{1-\nu}}. \tag{8}$$

The final term is obtained after making use of (7) and simplifying.

(ii) The slip associated with the trailing partial is no longer aligned with the loading direction. This means that in general there are three independent components of the slip at the crack tip $\boldsymbol{\delta}_{\rm tip}$ to be determined and a criterion of the form of (4) cannot be used. An approximate closed-form solution may be obtained by neglecting shear-tension coupling (i.e. setting $\delta_2 = 0$) and constraining the slip at the crack tip to coincide with the partial dislocation direction ($\delta_1 = \delta\cos\phi_B$, $\delta_3 = \delta\sin\phi_B$). As a result the crack tip singularity is not fully relaxed and residual SIFs $K_{\rm II(tip)}$ and $K_{\rm III(tip)}$ remain at the crack tip.

(iii) The ground level energy is no longer zero, but rather $\gamma_{\rm sf}$. Thus the barrier to dislocation emission is $\gamma_{\rm us} - \gamma_{\rm sf}$. This can be seen in Fig. 4 where the dashed line continuing from point III corresponds to the emission of the trailing partial. At the end of the emission the slip plane behind the resulting dissociated dislocation is undisturbed with zero energy penalty.

Given the factors listed above, reevaluating the J-integral along the far and slit contours gives,

$$J_{\rm far} = (1-\nu)K_{\rm II}^{*2}/2\mu = (1-\nu)K_{\rm II}^2/2\mu - \gamma_{\rm sf}, \tag{9}$$

$$J_{\rm slit} = \left[(1-\nu)K_{\rm II(tip)}^2 + K_{\rm III(tip)}^2\right]/2\mu + \Phi(\delta_{\rm tip}) - \Phi(\delta_{\rm sf}), \tag{10}$$

where $\delta_{\rm sf}$ is the slip discontinuity associated with the stacking fault left in the wake of the leading partial, such that $\Phi(\delta_{\rm sf}) = \gamma_{\rm sf}$.

Before the nucleation criterion can be obtained it is necessary to find $K_{\mathrm{II(tip)}}$ and $K_{\mathrm{III(tip)}}$. One equation for these quantities may be obtained from the solution for the effect of a slip distribution along the x-axis on the stress intensity factor,

$$K_{\mathrm{II(tip)}} - K^*_{\mathrm{II}} = \frac{\mu \cos\phi_B}{\sqrt{2\pi}(1-\nu)} \int_0^\infty \frac{1}{\sqrt{x_1}} \frac{d\delta(x_1)}{dx_1} dx_1, \tag{11}$$

$$K_{\mathrm{III(tip)}} = \frac{\mu \sin\phi_B}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{x_1}} \frac{d\delta(x_1)}{dx_1} dx_1. \tag{12}$$

Note that even for pure mode II loading, a mode III component exists at the tip. Rearranging (11) and (12) and equating between gives,

$$(1-\nu)\sin\phi_B \left[K_{\mathrm{II(tip)}} - K^*_{\mathrm{II}}\right] - \cos\phi_B K_{\mathrm{III(tip)}} = 0. \tag{13}$$

A second equation relating $K_{\mathrm{II(tip)}}$ and $K_{\mathrm{III(tip)}}$ is obtained by requiring that the shear stress in the direction of the slip be bounded. The shear stresses ahead of the crack tip are $\sigma_{21} = K_{\mathrm{II(tip)}}/\sqrt{2\pi r}$ and $\sigma_{23} = K_{\mathrm{III(tip)}}/\sqrt{2\pi r}$. The resolved shear stress in the slip direction is $\tau = \sigma_{21}\cos\phi_B + \sigma_{23}\sin\phi_B$. The condition for bounded shear stress at the crack tip is then,

$$\tau|_{r\to 0} = \frac{1}{\sqrt{2\pi r}} \left[\cos\phi_B K_{\mathrm{II(tip)}} + \sin\phi_B K_{\mathrm{III(tip)}}\right] < \infty. \tag{14}$$

Since ϕ_B, $K_{\mathrm{II(tip)}}$ and $K_{\mathrm{III(tip)}}$ are not functions of r, this condition can only be satisfied if,

$$\cos\phi_B K_{\mathrm{II(tip)}} + \sin\phi_B K_{\mathrm{III(tip)}} = 0. \tag{15}$$

This is the second equation for $K_{\mathrm{II(tip)}}$ and $K_{\mathrm{III(tip)}}$. By solving (13) and (15) together the tip SIFs are obtained,

$$K_{\mathrm{II(tip)}} = \frac{(1-\nu)\sin^2\phi_B}{\cos^2\phi_B + (1-\nu)\sin^2\phi_B} K^*_{\mathrm{II}}, \tag{16}$$

$$K_{\mathrm{III(tip)}} = -\frac{(1-\nu)\cos\phi_B \sin\phi_B}{\cos^2\phi_B + (1-\nu)\sin^2\phi_B} K^*_{\mathrm{II}}. \tag{17}$$

Substituting (16) and (17) into (10) and equating with (9), the critical SIF for the emission of the trailing partial and the generation of the dissociated dislocation is obtained, after some simplification, as

$$K^{\perp}_{\mathrm{II}} = \sqrt{2\mu \left[\frac{1}{1-\nu} + (1 - \frac{\gamma_{\mathrm{sf}}}{\gamma_{\mathrm{us}}})\tan^2\phi_B\right]\gamma_{\mathrm{us}}} = \lambda_{\mathrm{crit}} K^1_{\mathrm{II}}, \tag{18}$$

where

$$\lambda_{\mathrm{crit}} = \sqrt{1 + (1-\nu)(1 - \frac{\gamma_{\mathrm{sf}}}{\gamma_{\mathrm{us}}})\tan^2\phi_B}, \tag{19}$$

characterizes the additional load necessary to nucleate the trailing partial relative to the leading partial.

4 Emission of the twinning partial

Once the leading partial dislocation has been emitted (and before the emission of the trailing partial) there is a competition between two possible modes of deformation. The system will either nucleate the trailing partial of the dissociated dislocation as discussed in the previous section or it will nucleate another partial dislocation of the leading type on an adjacent plane and form a microtwin. Two quantify this competition, the critical SIF required to nucleate the twinning dislocation is computed and compared to the trailing partial nucleation criterion in (18). Since the nucleation of the microtwin involves the emission of a partial dislocation, exactly as in the previous cases, a Peierls analysis may be applied here as well. The derivation of the twinning criterion closely follows the derivation for the trailing partial in the previous section with two main exceptions:

(i) The twinning partial dislocation has the same orientation as the leading partial, so $\phi_B = \phi_A = 0$.
(ii) The maximum barrier which must be overcome to emit the twinning partial is not $\gamma_{\rm us} - \gamma_{\rm sf}$ but rather $\gamma_{\rm ut} - \gamma_{\rm sf}$. The physical significance of $\gamma_{\rm ut}$ is explained in Fig. 4 which shows the interplanar potential $\Psi(\Delta)$. As noted earlier, the first part of the curve from I to III corresponds to the emission of the leading partial and the dashed line to the emission of the trailing partial. Consider now instead the case where another leading partial dislocation is emitted on the plane above the plane with the stacking fault. This occurs along the curve from III to V. The energy barrier for this emission $\gamma_{\rm ut} - \gamma_{\rm sf}$ depends on the new maximum $\gamma_{\rm ut}$ (point IV) which is different from $\gamma_{\rm us}$. The parameter $\gamma_{\rm ut}$ is the maximum energy encountered when in a crystal containing an intrinsic stacking fault the part of the crystal one layer above the stacking fault is rigidly displaced along the twinning partial direction. This is demonstrated by atomic arrangements IV and V. After the emission of the second partial a microtwin is formed. An intrinsic stacking fault exists between the two partials and an extrinsic stacking fault (point V) exists behind the second partial. The energy of the extrinsic fault is $2\gamma_{\rm t}^*$ where $\gamma_{\rm t}^*$ is nearly equal to the twin boundary energy $\gamma_{\rm t}$ except for distant-neighbor interaction effects.

Given the above, the critical mode II SIF for nucleating the twinning dislocation $K_{\rm II}^{\rm T}$ is obtained from (18) by simply setting $\phi_B = 0$ and replacing $\gamma_{\rm us}$ with $\gamma_{\rm ut}$, so that

$$K_{\rm II}^{\rm T} = \sqrt{\frac{2\mu\gamma_{\rm ut}}{1-\nu}} = K_{\rm II}^{1}\sqrt{\frac{\gamma_{\rm ut}}{\gamma_{\rm us}}}. \tag{20}$$

Comparing this expression to (5) it is clear that $\gamma_{\rm ut}$ plays a similar role for twinning as $\gamma_{\rm us}$ does for dislocation emission. The analysis that follows demonstrates that the unstable twinning energy plays a key role in determining the propensity of a material to twin.

5 Twinning criterion

In the previous sections, the critical mode II SIFs for nucleating the leading partial dislocation $K_{\mathrm{II}}^{\mathrm{l}}$ (5), the trailing partial dislocation $K_{\mathrm{II}}^{\perp}$ (18) and the twinning partial dislocation $K_{\mathrm{II}}^{\mathrm{T}}$ (20) were obtained. The *twinning tendency* of a material T is defined as the ratio between the SIF for dislocation emission and the SIF for twinning,

$$T = \frac{K_{\mathrm{II}}^{\perp}}{K_{\mathrm{II}}^{\mathrm{T}}} = \lambda_{\mathrm{crit}} \sqrt{\frac{\gamma_{\mathrm{us}}}{\gamma_{\mathrm{ut}}}}. \tag{21}$$

A material will emit a dislocation before twinning if $T < 1$ and will twin first if $T > 1$. This is the twinning criterion. The twinning tendency expression in (21) lends itself to a clear physical explanation. T is affected by two factors: λ_{crit}, which characterizes the additional load necessary to nucleate the trailing partial relative to the leading partial, and $\gamma_{\mathrm{us}}/\gamma_{\mathrm{ut}}$, the ratio of the energy barriers for dislocation emission and DT. For cases where spontaneous emission of the trailing partial does not occur, λ_{crit} will be larger than one. The larger the value of λ_{crit} the larger that of T, signifying that DT is becoming more favorable due to the difficulty in nucleating the trailing partial. The ratio $\gamma_{\mathrm{us}}/\gamma_{\mathrm{ut}}$ will normally be less than one, so this factor tends to reduce T. In particular, the larger the value of γ_{ut} the more difficult it is for the material to twin. For cases where spontaneous emission of the trailing partial occurs, λ_{crit} will be less than one and hence T will also be less than one. The twinning criterion will thus predict dislocation emission in this case, consistent with the behavior of the system.

References

[DB83] Daw, M.S., Baskes, M.I.: Semiempirical, quantum-mechanical calculations of hydrogen embrittlement in metals. Phys. Rev. Lett., **50**, 1285–1288 (1983)

[EA93] Ercolessi, F., Adams, J.: Interatomic potentials from 1st-principles calculations – the force-matching method. Europhys. Lett., **26**, 583–588 (1993)

[Ric68] Rice, J.R.: A path independent integral and the approximate analysis of strain concentration by notches and cracks. J. of Appl. Mech., **35**, 379–386 (1968)

[Ric92] Rice, J.R.: Dislocation nucleation from a crack tip: an analysis based on the Peierls concept. J. Mech. Phys. Solids, **40**, 239–271 (1992)

[TH03] Tadmor, E.B., Hai, S.: A Peierls criterion for the onset of deformation twinning at a crack tip. J. Mech. Phys. Solids, **51**, 765–793 (2003)

[Vit68] Vitek, V.: Intrinsic stacking faults in body-centered cubic crystals. Phil. Mag., **18**, 773–786 (1968).

Part III

Physics/Chemistry/Fluid Dynamics/Biology

Simulation of Transport in Partially Miscible Binary Fluids: Combination of Semigrandcanonical Monte Carlo and Molecular Dynamics Methods

Kurt Binder, Subir K. Das, Jürgen Horbach, Marcus Müller, Richard Vink, and Peter Virnau

Institut für Physik, Johannes Gutenberg–Universität Mainz, Staudinger Weg 7, D–55099 Mainz, Germany, Kurt.Binder@uni-mainz.de

Binary Fluids that exhibit a miscibility gap are ubiquitous in nature (glass melts, polymer solutions and blends, mixtures of molten metals, etc.) and exhibit a delicate interplay between static and dynamic properties. This is exemplified for a simple model system, the symmetrical AB Lennard–Jones mixture. It is shown how semigrandcanonical Monte Carlo methods, that include A$\longrightarrow$B (B$\longrightarrow$A) identity switches as Monte Carlo moves, can yield the phase diagram, the interfacial tension between coexisting phases, and various pair correlation functions and structure factors. In addition to the build–up of long–ranged concentration correlations near the critical point, unmixing is also accompanied by the build–up of subtle structural features on very small length scales (less than the Lennard–Jones diameters).

System configurations, that are well equilibrated by these Monte Carlo methods, can be favorably used as initial states for Molecular Dynamics runs in the microcanonical ensemble. In this way it is possible to study the behavior of transport coefficients as well as dynamic correlation functions precisely along the coexistence curve, as well as for other paths in the one–phase region of the phase diagram.

1 Introduction

It has been a longstanding problem of physics and chemistry (and of materials science) to clarify the relationship between dynamic properties of fluid or solid matter, and the phase transitions which these systems undergo, in particular, phase separation. Typical materials are multicomponent metallic alloys, oxide mixtures, polymer solutions, and polymer blends [Kos01]. Phase separation in a polymer mixture is a multiscale problem — the polymer coil exhibits nontrivial structure from the scale of the length of covalent bonds (1 Å) to the scale of the gyration radius (100 Å), and in the critical region near the critical point of unmixing the correlation length of concentration fluctuations may be as large as 1000 Å [Bin94]. Also the dynam-

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

ics extends over many decades in time, from 10^{-13} seconds to macroscopic times [Bin95].

Sometimes phase separation competes with crystallization, for instance in the mixture of molten silicon dioxide and aluminium oxide. Again this is a multiscale problem: in the fluid there is nontrivial structure from the scale of covalent bonds to the medium range order that characterizes the continuous random network of this glassformer already in the melt far away from any phase transition [BK99]. Near the unmixing critical point we have again a large correlation length of concentration fluctuations. Also in this example the dynamics extends from subpicoseconds to macroscopic times, due to the proximity of the glass transition [Jae86]. This list of examples could be easily continued.

Dealing with chemically realistic models of such systems is still a major unsolved challenge. In this review we hence describe the state of the art of what can be done today by presenting a simple model system, the symmetrical binary AB Lennard–Jones mixture, and we will explain the appropriate methods of study. Then a selection of both static properties (phase diagram, interfacial tension, various radial distribution functions and structure factors that convey information on the unmixing) and quantities characterizing the dynamics (coherent and incoherent intermediate dynamic structure factors, wavenumber dependent relaxation times, selfdiffusion and interdiffusion constants, viscosity) will be defined and discussed. Emphasis always will be on the methodological aspects of such simulations.

2 A simple model: the symmetrical Lennard–Jones mixture

Following [DHB03] we choose a fluid of point particles located in continuous space (in a box of finite volume $V = L^3$ with periodic boundary conditions) and interacting with the potential

$$u(r_{ij}) = 4\epsilon_{\alpha\beta}\left[(\sigma_{\alpha\beta}/r_{ij})^{12} - (\sigma_{\alpha\beta}/r_{ij})^{6}\right], \quad r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|, \quad \alpha, \beta \in \mathrm{A}, \mathrm{B}\,. \tag{1}$$

The Lennard–Jones parameters $\{\epsilon_{\alpha\beta}, \sigma_{\alpha\beta}\}$ are chosen as follows

$$\sigma_{\mathrm{AA}} = \sigma_{\mathrm{AB}} = \sigma_{\mathrm{BB}} = \sigma, \qquad \epsilon_{\mathrm{AA}} = \epsilon_{\mathrm{BB}} = \epsilon\,. \tag{2}$$

Eq. (2) defines precisely what is meant by "symmetric mixture". In the following, units of length, temperature, and energy are chosen by requiring that $\sigma = 1$, $\epsilon = 1$, and Boltzmann's constant $k_{\mathrm{B}} = 1$. In the Molecular Dynamics (MD) part, also the masses are chosen as $m_{\mathrm{A}} = m_{\mathrm{B}} = m = 1$. In the Monte Carlo (MC) part, masses cancel out since we work with classical statistical mechanics (see e.g. [BC96] for a general introduction into MC and MD).

If the remaining energy parameter ϵ_{AB} would also be chosen as $\epsilon_{\mathrm{AB}} = \epsilon\ (= 1)$ one would have the trivial problem of an ideal mixture. So $\delta = \epsilon_{\mathrm{AB}}/\epsilon = 0.5 < 1$ is chosen: this leads to unmixing, because less energy is won for AB pairs rather than for AA and BB pairs (at the preferred distance $2^{1/6}\sigma$). As usual, the potential is truncated and shifted to zero at $r_{ij} = 2.5\sigma$ [BC96]. In order to stay away from the

gas–liquid transition as well as from crystallization, a density $\rho = 1$ is chosen. This ensures that the system stays in the fluid phase for the temperature range of interest, $1 \leq T \leq 2$ (note that the critical temperature of unmixing is $T_c \approx 1.638$, see below).

3 General comments on simulation methods

A first equilibration run is done with MC in the $N_A N_B V T$ ensemble for $N_A = N_B$ with standard trial displacements of particle coordinates in the range $[-\sigma/20,+\sigma/20]$, running for 100000 Monte Carlo steps (MCS) per particle. Then the semigrandcanonical MC algorithm is switched on: at the end of each displacement step per particle an attempted switch of $N/10$ randomly chosen particles is made, A→B or B→A (N denotes the total number of particles). These moves are accepted or rejected according to the standard Metropolis algorithm [BC96], but now both the energy change ΔE and the change of the chemical potential difference $\pm\Delta\mu$ needs to be taken into account in the Boltzmann factor. The independently fixed variables of the semigrandcanonical ensemble are total particle number N, volume V, temperature T, and chemical potential difference $\Delta\mu$, while the relative concentrations $x_A = N_A/(N_A + N_B) = N_A/N$, $x_B = N_B/N$ $(= 1 - x_A)$ are fluctuating variables. Their average values $< x_A >$, $< x_B >$ are an output rather than an input to the simulation! The symmetry between A and B in our model dictates that for the choice $\Delta\mu = 0$ we have $< x_B >= 1/2$ for $T > T_c$, while for $T < T_c$ one obtains states at the coexistence curve. The location of the coexistence curve shows up via peaks in the distribution function of the relative concentration $P(x_A)$, see Fig. 1.

In principle, for finite systems there is no symmetry breaking possible in full thermal equilibrium [BH02]. So for $T < T_c$, one should always observe two peaks, one at $x_A^{\mathrm{coex}(2)}$ and the other at $x_A^{\mathrm{coex}(1)}$, corresponding to the two branches of the coexistence curve. In practice, the two branches are separated at low temperatures by a huge free energy barrier. Therefore, with 400000 MCS "ergodicity breaking" [BH02] occurred for $T = 1.2$ and $T = 1.4$, i.e. only the A rich phase but not the B rich phase was observed. For $T = 1.6$, however both phases could be observed.

One can avoid "ergodicity breaking" by applying biased sampling techniques such as "successive umbrella sampling" [VM03] which then also yields quantitative information on the mentioned free energy barrier. In fact, one then can extract information on the surface tension between coexisting A–rich and B–rich phases from such a calculation, applying the formula [Bin82]

$$\frac{f_{\mathrm{int}}}{k_B T} = \lim_{L\to\infty} \frac{1}{2L^2} \ln\{P^{(\max)}(x_A)/P^{(\min)}(x_A)\} \; . \tag{3}$$

Fig. 2 shows the result of such a calculation for $N = 800$. The MC data are indeed compatible with the expected critical vanishing of f_{int} at T_c, $f_{\mathrm{int}} \propto (T_c - T)^{2\nu}$ where $\nu \approx 0.63$ [Bin82].

Fig. 3 shows the phase diagram of the symmetrical binary Lennard–Jones mixture. While the broken curve is a guide to the eye for the coexistence curve separating

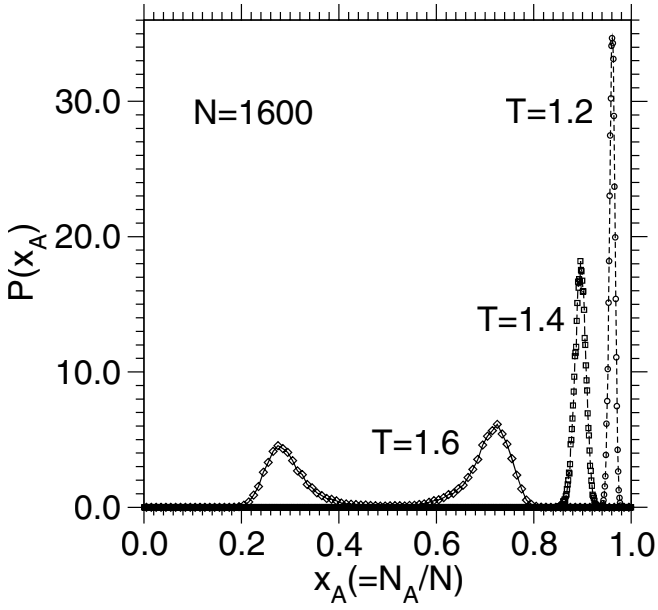


Fig. 1. Distribution function $P(x_{\rm A})$ for the relative concentration $x_{\rm A}$ of A particles for $N = 1600$ at $\Delta\mu = 0$ and the three temperatures indicated. From [DHB03].

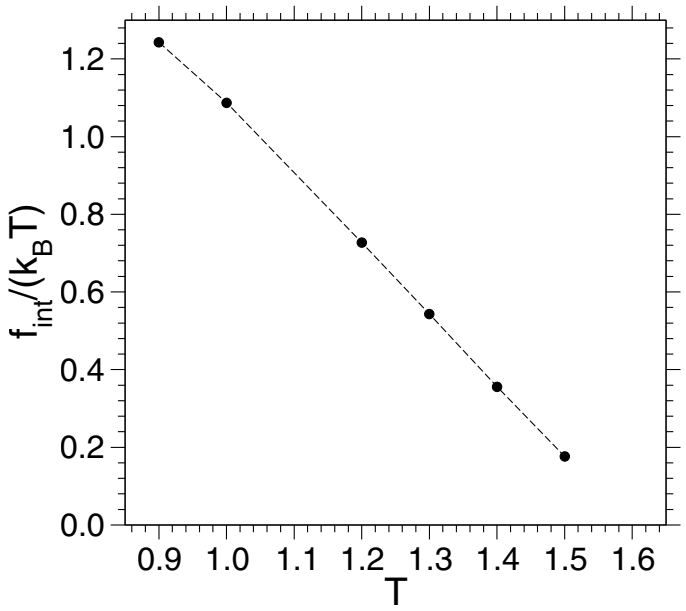


Fig. 2. Surface tension $f_{\rm int}/(k_BT)$ of the symmetrical binary Lennard–Jones mixture plotted vs. temperature.

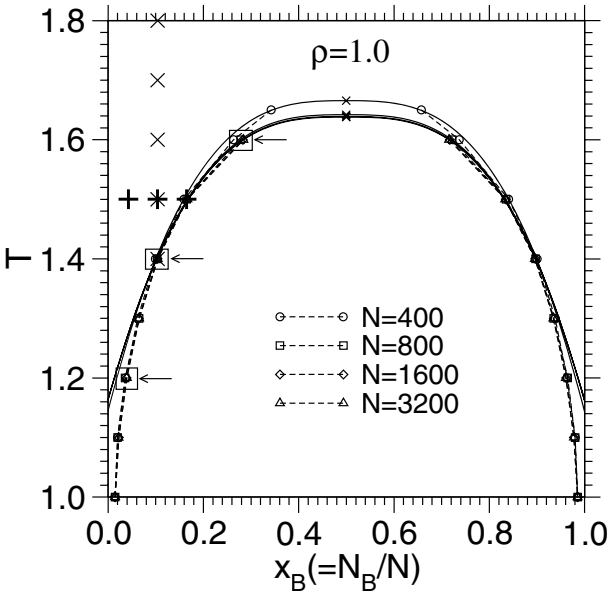


Fig. 3. Phase diagram of the symmetrical binary Lennard–Jones mixture, including results for four choices of N, as indicated. Crosses at $x_{\rm B} = 0.10375$, plus symbols at $T = 1.5$ and squares with arrows indicate paths in the $(T, x_{\rm B})$ plane for which the static and dynamical behavior was studied. From [DHB03].

the two–phase region (inside of the curve) from the one–phase region (outside of it), the full curves are phenomenological fits to the expected power law [BH02, BC96],

$$x_{\rm B}^{\rm crit} \pm x_{\rm B}^{\rm coex} = \hat{B}(1 - T/T_{\rm c})^{\beta}, \qquad \beta \approx 0.325 \quad , \tag{4}$$

where $\hat{B}$ is a so–called "critical amplitude" and of course, $x_{\rm B}^{\rm crit} = 1/2$, due to the symmetry of the model. From Fig. 3 it is clear that $N = 400$ is too small to allow reliable estimates, but for $N \geq 800$ the finite size effects are small, and the analysis of the MC data with Eq. (4) yields $T_{\rm c} = 1.638 \pm 0.005$. If higher accuracy were desired, a finite size scaling analysis [BC96, BH02] needed to be performed.

Due to the symmetry between A and B, "ergodicity breaking" is no problem at all here for locating the coexistence curve: one may normalize $P(x_{\rm A})$ such that

$$\int_{1/2}^{1} P(x_{\rm A}) dx_{\rm A} = 1, \qquad x_{\rm A}^{\rm coex(2)} = \int_{1/2}^{1} x_{\rm A} P(x_{\rm A}) dx_{\rm A} = 1 - x_{\rm A}^{\rm coex(1)} \quad . \tag{5}$$

If there were no symmetry between A and B, $x_{\rm A}^{\rm coex(1)}$ and $x_{\rm A}^{\rm coex(2)}$ would be no longer related via Eq. (5). In this case the "successive umbrella sampling" would allow to locate both concentrations $x_{\rm A}^{\rm coex(1)}$ and $x_{\rm A}^{\rm coex(2)}$ by the "equal weight rule" [BC96, BH02].

Five independent runs were made for $N = 1600$, storing ten statistically independent states, whose concentration is as close to $x_{\rm A}^{{\rm coex}(2)}$ as possible (note that $x_{\rm A}$ is quantized in steps of $1/N$). These 50 configurations (at each T) were used as starting states for MD runs (in the microcanonical NVE ensemble where the energy is conserved [BC96]), using the "velocity Verlet algorithm" [BC96] as time integrator, with a time step $\delta t = 0.01$, in units of the standard LJ time unit [BC96] $t_0 = (m\sigma^2/48\epsilon)^{1/2}$. In order to work in the one–phase region, one can "heat up" the configurations gotten from MC at fixed $x_{\rm B}$, and equilibrate the system by MD (the desired temperature is obtained by coupling the system to a stochastic heat bath [DHB03]).

It turns out that $N = 1600$ particles are enough to locate the coexistencence curve rather precisely, the critical temperature is estimated with an accuracy better than 0.5%. If needed, much better accuracy could be achieved by a careful finite size scaling analysis [BC96, BH02]. In the following, the physical properties of the system along three paths will be described (Fig. 3): along the coexistence curve; at constant composition $x_{\rm B} = 0.10375$, changing the temperature; and at the isotherm $T = 1.5$, crossing the one–phase region up to the coexistence curve.

4 Static properties

A static property of central interest in all simulations of fluids are radial distribution functions [BC96]. In a binary system three types of distribution functions have to be distinguished, $g_{\rm AA}(r)$, $g_{\rm AB}(r)$, and $g_{\rm BB}(r)$. These are defined as

$$g_{\alpha\beta}(r) = \frac{N}{\rho N_{\rm A} N_{\rm B}} \left\langle \sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\beta}{}' \frac{\delta(r - |\mathbf{r}_i(0) - \mathbf{r}_j(0)|)}{4\pi r^2} \right\rangle \quad . \tag{6}$$

As an example, Fig. 4 shows these partial radial distribution functions for a state at the coexistence curve at $T = 1.2$. Although the model is symmetric between A and B, $g_{\rm AA} \neq g_{\rm BB}$. This happens, because along the A–rich branch of the coexistence curve, the concentration of B particles is less than that of A particles. When one increases the temperature, this difference gets smaller and ultimately must vanish at the critical point.

Of particular interest are also the static structure factors $S_{\alpha\beta}(q)$, since these quantities in principle are experimentally accessible in elastic scattering experiments (e.g. neutron scattering with momentum transfer $\hbar q$). Again one has to distinguish three partial structure factors, defined according to

$$S_{\alpha\beta}(q) = \frac{f_{\alpha\beta}}{N} \sum_{k=1}^{N_\alpha} \sum_{l=1}^{N_\beta} \langle \exp[i\mathbf{q} \cdot (\mathbf{r}_k - \mathbf{r}_l)] \rangle \tag{7}$$

with $f_{\alpha\beta} = 1$ for $\alpha = \beta$ and $f_{\alpha\beta} = 0.5$ for $\alpha \neq \beta$. From the partial structure factors one can form various combinations of physical interest, such as the number density

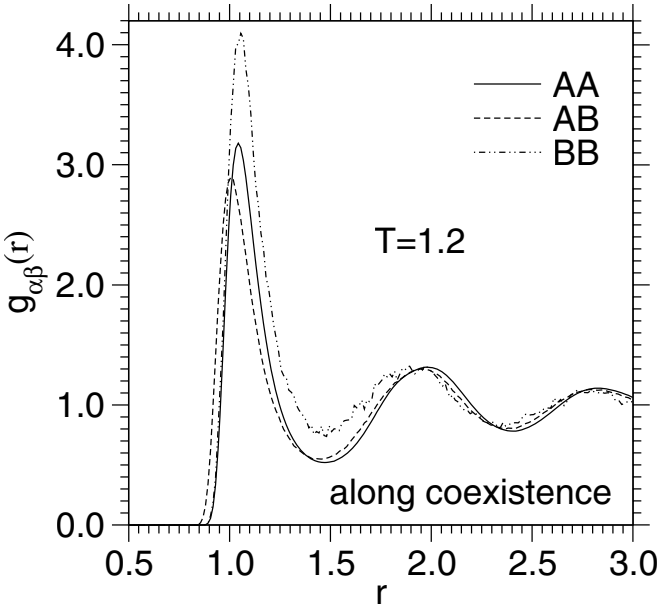


Fig. 4. Radial distribution function $g_{\rm AA}(r)$ (solid line), $g_{\rm AB}(r)$ (dashed line), and $g_{\rm BB}(r)$ (dashed–dotted line) plotted vs. r, for a state at the coexistence curve at $T = 1.2$. From [DHB03].

structure factor $S_{\rm nn}(q)$,

$$S_{\rm nn}(q) = S_{\rm AA}(q) + 2S_{\rm AB}(q) + S_{\rm BB}(q) \quad . \tag{8}$$

$S_{\rm nn}(q)$ would show up in scattering of radiation that does not distinguish between the particles. Fig. 5 shows that $S_{\rm nn}(q)$ appears to behave as any ordinary (almost incompressible) fluid {note that $S_{\rm nn}(q \to 0)$ is related to the isothermal compressibility [BC96]}. There is no sign in this quantity that a critical point of unmixing is approached. However, there is another combination [Zim79],

$$S_{\rm cc}(q) = x_{\rm B}^2 S_{\rm AA}(q) + x_{\rm A}^2 S_{\rm BB}(q) - 2x_{\rm A}x_{\rm B}S_{\rm AB}(q) \; , \tag{9}$$

which is basically the Fourier transform of the correlation function $g_{\rm cc}(r)$ of concentration fluctuations $\{g_{\rm cc}(r) = x_{\rm A}^2 x_{\rm B}^2 \, [g_{\rm AA}(r) + g_{\rm BB}(r) - 2g_{\rm AB}(r)]\}$.

It is remarkable that $g_{\rm cc}(r)$ exhibits a negative peak around $r \approx 0.9$ (Fig. 6). This indicates an anti–correlation in the concentration density fluctuations which stems from the fact that distances around $r = 0.9$ are avoided by nearest A–A and B–B neighbors but are typical for nearest A–B neighbors. While $g_{\rm cc}(r)$ is almost zero at low temperatures and in addition very short–ranged (Fig. 6), it develops a stronger peak and a more long–ranged slow decay as one approaches criticality. One expects $g_{\rm cc} \propto r^{-1} \exp(-r/\xi)$, ξ being the correlation length of concentration fluctuations. The maximum value of $S_{\rm cc}(q)$ occurs for $q = 0$. This maximum could never be

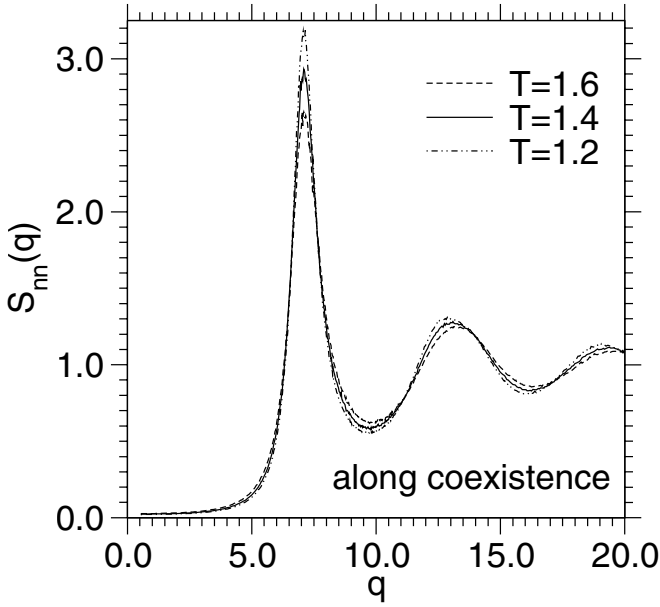


Fig. 5. Structure factor $S_{\rm nn}(q)$ plotted vs. q, for the three indicated temperatures, choosing states along the coexistence curve. From [DHB03].

directly estimated from a microcanonical MD simulation, where the number of A particles as well as the number of B particles is fixed and thus $S_{\rm cc}(q=0)=0$ (in this case the nonzero maximum rather would be defined via the limit $S_{\rm cc}(q \to 0^+)$). However, the semigrandcanonical MC calculation allows to estimate this quantity via a simple fluctuation relation,

$$S(q=0) = k_B T \chi = N \left\{ \int_{1/2}^{1} x_{\rm A}^2 P(x_{\rm A}) dx_{\rm A} - [x_{\rm A}^{\rm coex(1)}]^2 \right\} \quad . \tag{10}$$

The values of $k_B T \chi$ as obtained by Eq. (10) are shown in Fig. 6 as crosses at $q=0$.

Interestingly, there is also a nonzero cross correlation between concentration and density fluctuations. This is measured via the function $g_{\rm nc}(r) = x_{\rm A} x_{\rm B} [x_{\rm A} g_{\rm AA}(r) - x_{\rm B} g_{\rm BB}(r) + (x_{\rm B} - x_{\rm A}) g_{\rm AB}(r)]$ or the corresponding structure factor $S_{\rm nc}(q)$,

$$S_{\rm nc}(q) = x_{\rm B} S_{\rm AA}(q) - x_{\rm A} S_{\rm BB}(q) + (x_{\rm B} - x_{\rm A}) S_{\rm AB}(q) \quad . \tag{11}$$

If one would work exactly at $x_{\rm A} = x_{\rm B} = 1/2$ for $T > T_{\rm c}$, symmetry would require that $S_{\rm nc}(q) \equiv 0$. So it is plausible that along the coexistence curve this quantity is still rather small (Fig. 7).

The same structure factors have also been studied for the other paths included in the phase diagram, Fig. 3. As an example, Fig. 8 shows $S_{\rm cc}(q)$ for an isotherm at $T = 1.5$. Further simulation data can be found in [DHB03].

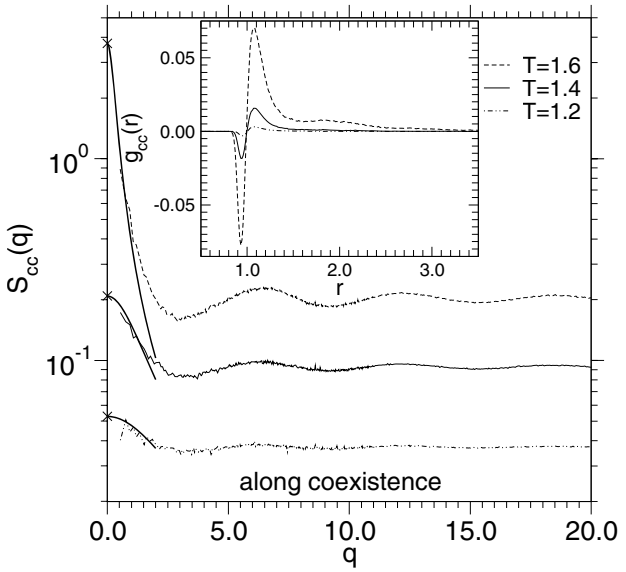


Fig. 6. Concentration fluctuation structure factor $S_{\rm cc}(q)$ plotted vs. q, for three temperatures as indicated, choosing states along the coexistence curve. Note the logarithmic scale of the ordinate. Crosses at the ordinate represent the "susceptibility" $k_BT\chi$ (see text), while the bold solid lines are fits to the Ornstein–Zernike equation $S_{\rm cc}(q) = k_BT\chi/[1 + q^2\xi^2]$, which should hold at small q. Inset shows the correlation function $g_{\rm cc}(r)$. From [DHB03].

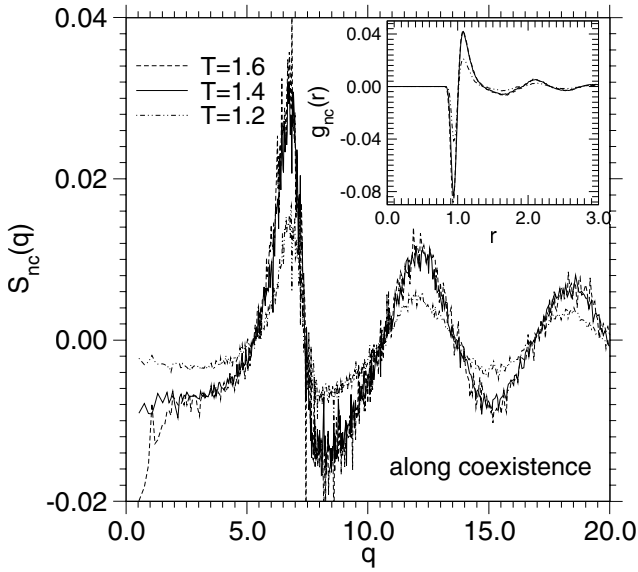


Fig. 7. Cross correlation structure factor $S_{\rm nc}(q)$ plotted vs. q, for three temperatures as indicated, choosing states along the coexistence curve. The inset shows the correlation function $g_{\rm nc}(r)$. From [DHB03].

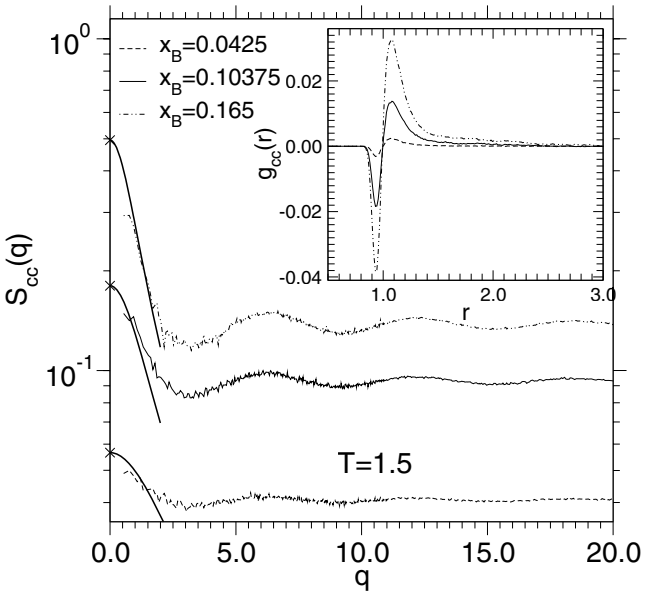


Fig. 8. Concentration fluctuation structure factor $S_{\rm cc}(q)$ plotted vs. q, at $T = 1.5$ and three concentrations as shown. The inset shows $g_{\rm cc}(r)$. For further explanations see Fig. 6. From [DHB03].

5 Dynamic properties

A quantity that contains information on how the various kinds of particles move, and that is experimentally accessible e.g. via incoherent inelastic scattering of neutrons, is the incoherent intermediate scattering function $F_{\rm s}^{(\alpha)}(q,t)$,

$$F_{\rm s}^{(\alpha)}(q,t) = \frac{1}{N_\alpha} \sum_{i \in \alpha} \langle \exp\{-i\mathbf{q} \cdot [\mathbf{r}_i(0) - \mathbf{r}_i(t)]\} \rangle \,, \qquad \alpha \in (A,B) \quad . \tag{12}$$

Fig. 9 shows this quantity for the A particles, including three states along the coexistence curve. The decay of this function with time t obviously becomes very slow when q becomes small. This observation simply describes the "hydrodynamic slowing down": the relaxation times diverge as expected from diffusive behavior. A simple argument to see this assumes the displacements $\mathbf{r}_i(0) - \mathbf{r}_i(t)$ to be Gaussian distributed at long enough times, which yields $F_{\rm s}^{(\alpha)}(q,t) = \exp[-q^2 g_\alpha(t)/6]$, where $g_\alpha(t)$ is the mean square displacement of species α,

$$g_\alpha(t) = \left\langle [\mathbf{r}_{i,\alpha}(0) - \mathbf{r}_{i,\alpha}(t)]^2 \right\rangle \quad . \tag{13}$$

Using now the Einstein relation, $g_\alpha(t) = 6D_\alpha t$ for $t \to \infty$, one notes that $F_{\rm s}^{(\alpha)}(q,t) \propto \exp[-t/\tau_\alpha(q)]$ with $\tau_\alpha(q) = (D_\alpha q^2)^{-1}$. Fig. 10 shows that this interpretation indeed is compatible with the MD results for $q \to 0$. This hydrodynamic

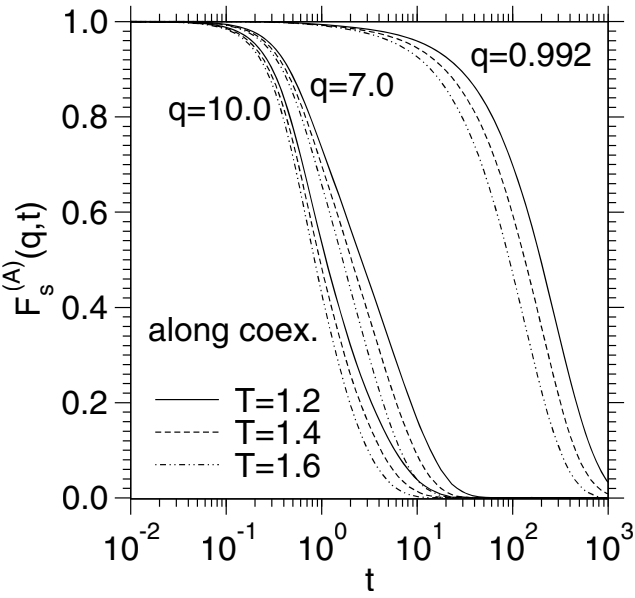


Fig. 9. Incoherent intermediate scattering function of A particles plotted vs. time (note the logarithmic scale of time) for three temperatures and concentrations, along the A–rich part of the coexistence curve, $T = 1.2$ (solid lines), $T = 1.4$ (dashed lines), and $T = 1.6$ (dashed–dotted lines). Each panel shows $F_{\rm s}^{(\rm A)}(q,t)$ for 3 values of q, namely $q = 0.992, 7, 10$ (from right to left). From [DHB03].

slowing down is a general consequence of the fact that in the microcanonical MD simulation the concentrations $x_{\rm A} = N_{\rm A}/N$, $x_{\rm B} = N_{\rm B}/N$ are conserved quantities: Therefore, long–wavelength concentration fluctuations, if they were present in the system, would take a very long time to equilibrate. Clearly, this equilibration problem is avoided by the equilibration with Monte Carlo methods in the semigrand-canonical ensemble.

Fig. 11 presents data on the coherent intermediate scattering function $F_{\alpha\beta}(q,t) \equiv S_{\alpha\beta}(q,t)/S_{\alpha\beta}(q,0)$ where

$$S_{\alpha\beta}(q,t) = \frac{f_{\alpha\beta}}{N} \sum_{i\in\alpha} \sum_{j\in\beta} \langle \exp\{-i\mathbf{q}\cdot[\mathbf{r}_i(t) - \mathbf{r}_j(0)]\}\rangle \tag{14}$$

with $f_{\alpha\beta} = 1$ for $\alpha = \beta$ and $f_{\alpha\beta} = 0.5$ for $\alpha \neq \beta$. From $F_{\rm AA}(q,t)$ {Fig. 11} one can clearly recognize that several relaxation processes contribute, since the decay occurs in several steps. In contrast, in the concentration correlation function $F_{\rm cc}(q,t)$ only a single step (representing the slowest decay) is found (Fig. 12). For the cases that were studied, the relaxation time describing this "'slow" decay is not yet much slower than the decay times $\tau_{\rm A}(q)$ and $\tau_{\rm B}(q)$ of the self–correlation functions, Figs. 9 and 10. However, taking data close to the critical point one expects that the collective

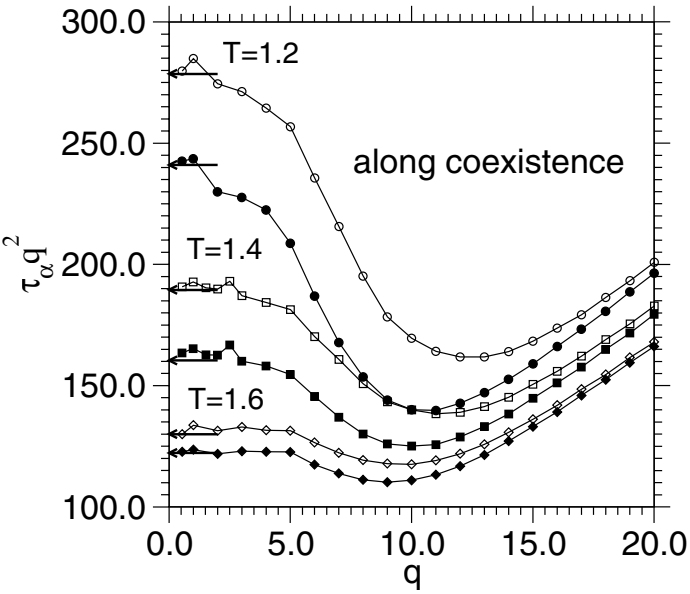


Fig. 10. Plot of the scaled relaxation time $\tau(q)q^2$ vs. q, for the A particles (open symbols), and the B particles (filled symbols), for three temperatures along the A–rich branch of the coexistence curve. The arrows indicate the values of the corresponding inverse selfdiffusion constants, as estimated from the Einstein relation. From [DHB03].

time scale seen in $F_{\rm cc}(q,t)$ should become very much larger than the single–particle timescales $\tau_{\rm A}(q)$ and $\tau_{\rm B}(q)$, which are related to selfdiffusion, as noted above. One does not expect any significant slowing down of selfdiffusion when one approaches the critical point [KBK82]. In contrast, $F_{\rm cc}(q,t)$ should pick up the decay described by the interdiffusion constant [KBR89]: in addition to the "hydrodynamic slowing down" shown by all relaxation times relating the diffusion because of the conservation laws for the concentration of particles [KM63], there is also "critical slowing down" [HH77], i.e. the interdiffusion constant vanishes as $T \to T_{\rm c}$, and this slow decay one expects to see in $F_{\rm cc}(q,t)$. However, for the choices of T presented this is not yet evident, since the onset of critical slowing down in this regime of temperature is to a large extent offset by a general slowing down of all transport phenomena as the temperature is lowered.

This ordinary slowing down (which often is described by simple Arrhenius laws for transport coefficients, e.g. $\ln(D_{\alpha\beta}) = -E_{\alpha\beta}/T + \text{const.}$, $E_{\alpha\beta}$ being an activation energy) can be studied conveniently when one follows a path at constant concentration in the one–phase region above the coexistence curve, such as the path at $x_{\rm B} = 0.10375$ included in Fig. 3. The corresponding results are shown in Fig. 13, where the selfdiffusion coefficients $D_{\rm A}$, $D_{\rm B}$ are shown together with the inverse shear viscosity η^{-1} and the interdiffusion coefficient $D_{\rm int}$, as extracted from Green–Kubo formulas [BY80]. These relations are

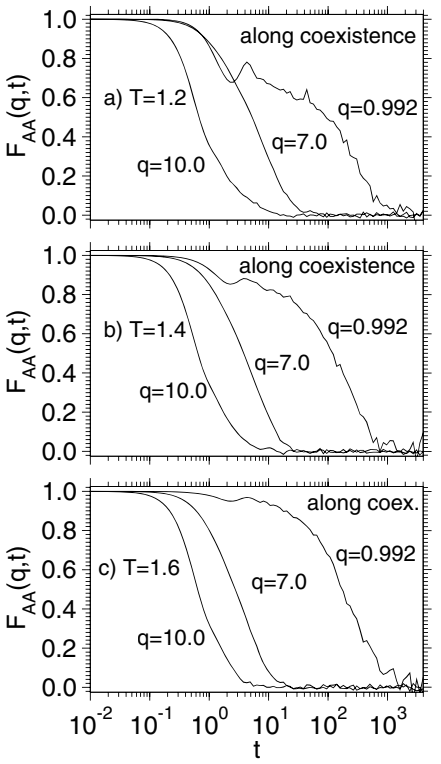


Fig. 11. Coherent intermediate scattering function $F_{\rm AA}(q,t)$ of A particles plotted vs. time (note the logarithmic scale of time). Three temperatures are shown at states along the coexistence curve: $T = 1.2$ (a), $T = 1.4$ (b), and $T = 1.6$ (c). Three wavenumbers are included as indicated. From [DHB03].

$$\eta = \frac{1}{Vk_BT}\int_0^\infty dt\ \langle\sigma_{xy}(0)\sigma_{xy}(t)\rangle \tag{15}$$

where $\sigma_{xy}(t)$ is an off–diagonal component of the pressure tensor defined by

$$\sigma_{xy}(t) = \sum_i \left(m_i v_{ix} v_{iy} + \frac{1}{2}\sum_{j\neq i} |x_i - x_j| F_y(|\mathbf{r}_i - \mathbf{r}_j|) \right), \tag{16}$$

$\mathbf{v}_i$ being the velocity of the i'th particle, and $\mathbf{F}(|\mathbf{r}_i - \mathbf{r}_j|)$ is the force acting on particle i due to the (pairwise) interaction with particle j. The sum $\sum_i$ in Eq. (16) extends over both kinds of particles, unlike Eq. (14). A similar relation can also be formulated for the interdiffusion coefficient, namely

$$D_{\rm int} = \frac{1}{NS_{\rm cc}(0)}\int_0^\infty dt\,\left\langle j_x^{\rm int}(0) j_x^{\rm int}(t)\right\rangle, \tag{17}$$

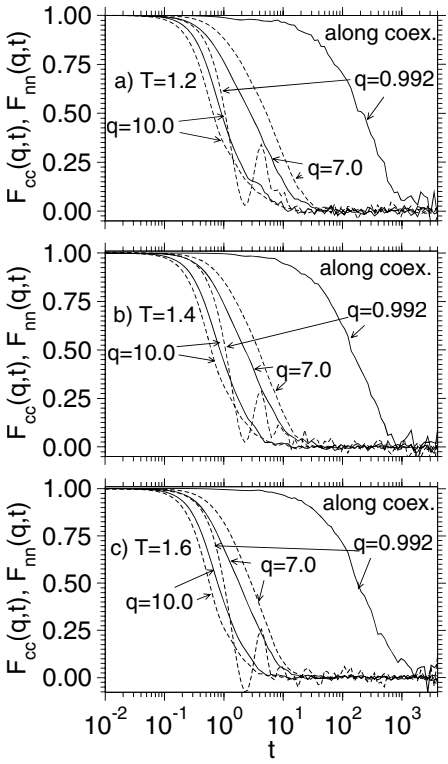


Fig. 12. Concentration–concentration correlation functions $F_{\rm cc}(q,t) = S_{\rm cc}(q,t)/S_{\rm cc}(q,0)$ plotted vs. time. Here $S_{\rm cc}(q,t)$ is the dynamic counterpart of the function $S_{\rm cc}(q)$ defined in Eq. (9). Three states along the coexistence are shown with $T = 1.2$ (a), $T = 1.4$ (b), and $T = 1.6$ (c). Curves for the same wavenumbers q as shown in Fig. 11 are included. From [DHB03].

with the interdiffusion current given by

$$\mathbf{j}_{\rm int}(t) = x_{\rm B} \sum_{i=1}^{N_{\rm A}} \mathbf{v}_i(t) - x_{\rm A} \sum_{i=1}^{N_{\rm B}} \mathbf{v}_i(t) \quad . \tag{18}$$

From Fig. 13 it is evident that all transport coefficients $D_{\rm A}$, $D_{\rm B}$, $D_{\rm int}$ and η^{-1} decrease as one lowers the temperature, but simple Arrhenius relations do not hold. In addition, we note that for the minority component the selfdiffusion constant always is a bit faster than for the majority component. The interdiffusion constant has a value which is hardly different from the selfdiffusion constants at high temperatures, but becomes appreciably smaller than the latter when one approaches the coexistence curve. Recall, as noted above, one expects $D_{\rm int} \to 0$ if the path $x_{\rm B} = x_{\rm B}^{\rm crit} = 1/2$ would be chosen so that one approaches the critical point, while for the strongly off–critical case considered in Fig. 13 $D_{\rm int} \neq 0$ at the coexistence curve, of course.

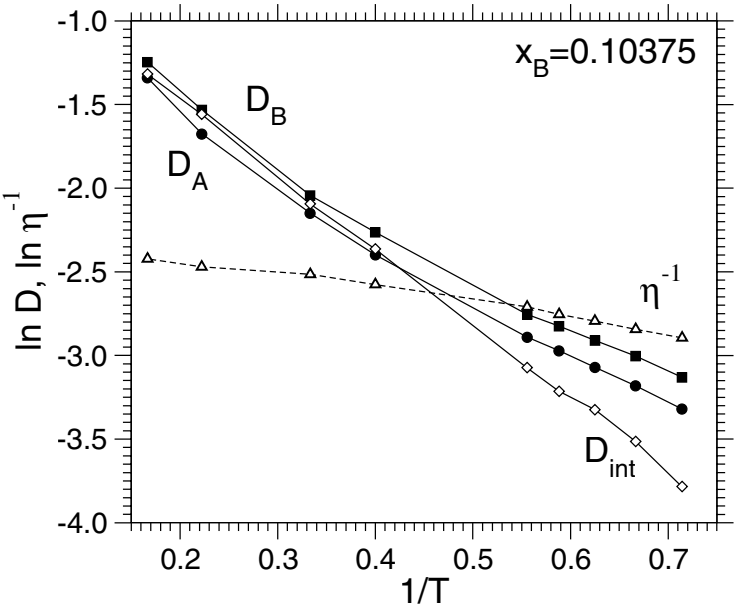


Fig. 13. Logarithm of the selfdiffusion constants D_{A}, D_{B}, the interdiffusion constant D_{int}, and the inverse viscosity η^{-1} plotted vs. inverse temperature for the mixture at constant concentration $x_{\mathrm{B}} = 0.10375$. From [DHB03].

The results for the viscosity allow a check of the Stokes–Einstein formula for the selfdiffusion constants

$$D_\alpha = \frac{k_B T}{2\pi\eta d_\alpha} \quad , \tag{19}$$

d_α being the (effective) diameter of the corresponding particle species $\alpha \in \{\mathrm{A}, \mathrm{B}\}$. If Eq. (19) were strictly correct, one would expect $d_{\mathrm{A}} \approx \sigma_{\mathrm{A}} = 1$ and $d_{\mathrm{B}} \approx \sigma_{\mathrm{B}} = 1$. It turns out that this would be a reasonable approximation for the A particles, while for the B particles the result rather is $d_{\mathrm{A}} \approx 0.8$ [DHB03]. Note that Eq. (19) is a widely used approximation, but it really lacks a good theoretical foundation: so this deviation of the simulation data from Eq. (19) is not really surprising.

6 Discussion and outlook

Even for the simple model system of a symmetrical Lennard–Jones mixture there remains a lot to be done: the critical region needs to be explored for much larger particle numbers N, in conjunction with a finite size scaling analysis, to probe the theoretical predicted critical singularities [HH77] of the various transport coefficients. Such a study has never been carried our before.

Then it would be very interesting to simulate “quenching experiments” where one e.g. suddenly changes the temperature (at fixed x_{A}, x_{B}) from a state above the

coexistence curve to a state below it, bringing the system into the two–phase region. Then the initially homogeneous state is unstable and will decay by either nucleation and growth (if the state point is close enough to the coexistence curve so that that the homogeneous state is metastable) or by "'spinodal decomposition" [Kos01]. Simulations of these processes are available for various models of solid mixtures (e.g. kinetic Ising models of rigid solid binary alloys) and in the framework of numerical work with suitable differential equations (e.g. the nonlinear Cahn–Hilliard equation [Kos01]), but atomistic simulations of these processes for liquid binary mixtures are not yet available. Such simulations are rather demanding, however, since the inhomogeneous two–phase structures that form are characterized by length scales much larger than the atomic diameters. Other interesting projects for this model (some of them are already in progress!) consider confinement effects in slit capillaries of nanoscopic size, nonlinear flow problems, etc.

However, the real challenge is to carry the strategy described in the present paper over to chemically realistic models of real materials. Note that there is nontrivial structure in the pair correlation functions (Figs. 6 and 7) on scales much less than the Lennard–Jones parameters (describing such effects for real materials will require very accurate interatomic potentials as an input, which often are not yet available), in addition to the structure on the scale of the correlation length ξ (which grows to large values near the critical point). Real binary systems never have strictly symmetric miscibility gaps, and then phase coexistence occurs (in the semigrandcanonical ensemble) along a nontrivial curve $\Delta\mu_{\rm coex}(T)$ in the $(\Delta\mu, T)$ plane. Mapping out the coexistence curve is a much harder task (though this can be done with the successive umbrella sampling technique [VM03]). One should also note that for asymmetric mixtures the A→B identity switch step may suffer from very low acceptance rates and suitable biasing schemes may need to be developed.

If all these problems could be solved, the present approach can yield very important input information for problems dealing with fluids on mesoscopic scales (where techniques such as Lattice–Boltzmann can be applied) or even macroscopic scales (where methods of computational fluid dynamics can be applied). The present approach therefore is potentially useful for applications in many fields (e.g. flow of oil plus water through porous rocks, flow of multiphase mixtures of polymers in the processing of plastic materials, flow of molten silicates in geophysical problems, etc.), but there remains still a long way to go until this will become a standard ingredient of multiscale simulation approaches.

Acknowledgments: One of us (S. K. D.) was supported by the Deutsche Forschungsgemeinschaft (DFG) und grant N$^{\rm o}$ Bi314-18 (SPP1120), another (J. H.) by DFG grant HO2231/2-1, another (M. M.) by a Heisenberg stipend, R. V. by SFB TR6 (A5). P. V. was partially supported by the Bundesministerium für Bildung und Forschung (BMBF) under grant N$^{\rm o}$ 03N6015.

References

[BC96] Binder, K., Ciccotti, G. (eds.): *Monte Carlo and Molecular dynamics of Condensed Matter Systems*, Italian Physical Society, Bologna (1996)

[BH02] Binder, K., Heermann, D. W.: *Monte Carlo Simulations in Statistical Physics. An Introduction, 4'th edition*, Springer, Berlin (2002)

[Bin82] Binder, K.: *Monte Carlo calculation of the surface tension for two– and three–dimensional lattice–gas models*, Phys. Rev. A **25**, 1699–1709 (1982)

[Bin94] Binder, K.: *Phase transitions in polymer blends and block copolymer melts: some recent developments*, Adv. Polymer Sci. **112**, 181–299 (1994)

[Bin95] Binder, K.: *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, Oxford University Press, New York (1995)

[BK99] Bach, H., Krause, D. (eds.): *Characterization of Structure and Properties of Glass*, Springer, Berlin (1999)

[BY80] Boon, J. P., Yip, S.: *Molecular Hydrodynamics*, McGraw Hill, New York (1980)

[DHB03] Das, S. K., Horbach J., Binder, K.: *Transport phenomena and microscopic structure in partially miscible binary fluids: A simulations study of the symmetrical Lennard–Jones mixture*, J. Chem. Phys. **119**, 1547–1558 (2003)

[HH77] Hohenberg, P. C., Halperin, B. I.: *Theory of dynamic critical phenomena*, Rev. Mod. Phys. **49**, 435–479 (1977)

[Jae86] Jäckle, J.: *Models of the glass transition*, Rep. Progr. Phys. **49**, 171–231 (1986)

[KBK82] Kutner, R., Binder, K., Kehr, K. W.: *Diffusion in concentrated lattice gases. II. Particles with attractive nearest-neighbor interaction on three-dimensional lattices*, Phys. Rev. B **26**, 2967–2980 (1982)

[KBR89] Kehr, K. W., Binder, K., Reulein, S. M.: *Mobility, interdiffusion, and tracer diffusion in lattice-gas models of two-component alloys*, Phys. Rev. B, **39**, 4891–4910 (1989)

[KM63] Martin, P., Kadanoff, L. P.: *Hydrodynamic equations and correlation functions*, Ann. Phys. (NY) **24**, 419–469 (1963)

[Kos01] Kostorz, G. (ed.): *Phase Transformations in Materials*, Wiley-VCH, Berlin (2001)

[VM03] Virnau, P., Müller, M.: *Successive umbrella sampling*, preprint (2003)

[Zim79] Ziman, J. M.: *Models of Disorder*, Cambridge University Press, Cambridge (1979)

Computer simulations of SiO_2 and GeO_2

Michael Hawlitzky, Jürgen Horbach, and Kurt Binder

Institut für Physik, Johannes Gutenberg–Universität Mainz, Staudinger Weg 7, D–55099 Mainz, hawlitzky@uni-mainz.de, horbach@uni-mainz.de

Classical Molecular Dynamics (MD) simulations are used to study structural and dynamic properties of amorphous germania (GeO_2) in comparison to those of silica (SiO_2). The total structure factor, as obtained from these simulations, is in very good agreement with that of neutron scattering experiments, both for germania and silica. The tetrahedral network structure in silica and germania leads to a prepeak in the structure factor that appears at slightly smaller wavenumbers in GeO_2 than in SiO_2. At high temperatures the diffusion constants are very similar in both systems whereas at low temperatures diffusion is significantly faster in germania than in silica. We also outline the strategy for the development of a potential for SiO_2–GeO_2 mixtures by means of Car–Parrinello MD.

1 Introduction

Amorphous SiO_2 is the prototype of a glassformer that forms a disordered tetrahedral network where the SiO_4 tetrahedra are connected via the oxygens at the corners. In the past 10 years many efforts have been undertaken to understand the static and dynamic properties of amorphous SiO_2(see [HK99, SSP01] and references therein) as well as mixtures of SiO_2 with other oxides such as Na_2O, Al_2O_3 (see e.g. [JKJ01, HKB02, LKH03, WHK03]) etc. by means of large scale MD simulations. It is surprising that similar studies have not been done for another glassformer that forms a tetrahedral network, GeO_2. Therefore, in the first step of our work we aim to close this gap and to give a detailed comparison of the differences between SiO_2 and GeO_2 with respect to structural and dynamic properties.

Our next step will be the investigation of mixtures of SiO_2 and GeO_2 which is particularly interesting since such mixtures are paradigms of real (technologically important) systems where the cations Si and Ge exhibit a very similar chemical ordering. Although it is of great technological importance, it is still not known whether in such a mixture the cationic species are distributed homogeneously or heterogeneously on a nanoscopic lengthscale. However, no accurate interaction potentials for GeO_2–SiO_2 mixtures are available in the literature. Our idea is to parametrise

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

such potentials by means of Car–Parrinello MD [CP85] which is an *ab initio* technique where the electronic degrees of freedom are explicitly taken into account in the framework of density functional theory.

The rest of the paper is organized as follows: In the next section we present the model potentials that we have used to simulate pure SiO_2 and GeO_2. In Sec. 3 we describe the strategy to derive a potential function for GeO_2–SiO_2 mixtures by means of Car–Parrinello MD. Then, we show in Sec. 4 the results of the MD simulation of pure SiO_2 and GeO_2 and we finally summarize the results.

2 Model Potentials for Silica and Germania

Model potentials for $\mathrm{SiO_2}$ and $\mathrm{GeO_2}$, that are based on *ab initio* calculations, have been proposed by van Beest, Kramer, and van Santen [BKS90] (BKS potential) and by Oeffner and Elliott [OE98] (OE potential), respectively. Both models describe the interactions between the atoms by pair potentials that have the same functional form given by

$$V_{\alpha\beta}(r) = \frac{q_\alpha q_\beta e^2}{4\pi\varepsilon_0 r} + a_{\alpha\beta}\exp\left(-\frac{r}{\rho_{\alpha\beta}}\right) - \frac{c_{\alpha\beta}}{r^6} \quad , \tag{1}$$

with r being the distance between a particle of type α and a particle of type β, where $\alpha, \beta \in \{\mathrm{Si}, \mathrm{Ge}, \mathrm{O}\}$. The partial charges q_α and the parameters $a_{\alpha\beta}$, $\rho_{\alpha\beta}$, and $c_{\alpha\beta}$ can be found in the original publications [BKS90, OE98]. e and ε_0 in the coulomb term denote the elementary charge and the permeability constant, respectively.

One may wonder why simple pair potentials can give an accurate description of the tetrahedral network structure in $\mathrm{SiO_2}$ and $\mathrm{GeO_2}$. One may expect that three–body terms have to be taken into account to stabilise respectively the intra–tetrahedral O–Si–O and O–Ge–O angles. But one has to be aware of the fact that by the interplay of three different pair potentials in the models for $\mathrm{SiO_2}$ and $\mathrm{GeO_2}$ many–body effects are taken into account effectively. In contrast, one–component systems such as silicon require the consideration of three–body terms in the potential function.

Fig. 1 shows some of the $V_{\alpha\beta}$ from the BKS and the OE model. We see that the Ge–O and the Si–O potentials exhibit minima at different locations and with a different depth. Note that the difference of the latter two potentials at large r is due to the different partial charges used for Ge and Si, $q_{\mathrm{Ge}} = 1.5$ and $q_{\mathrm{Si}} = 2.4$. As a consequence, the OE value for q_{O} is -0.75 as opposed to the BKS value $q_{\mathrm{O}} = -1.2$ which leads to very different O–O interaction potentials in both models (see Fig. 1). The means that the BKS and the OE model cannot be combined in a simple way to consider mixtures of $\mathrm{SiO_2}$ with $\mathrm{GeO_2}$ because this would require identical O–O interactions.

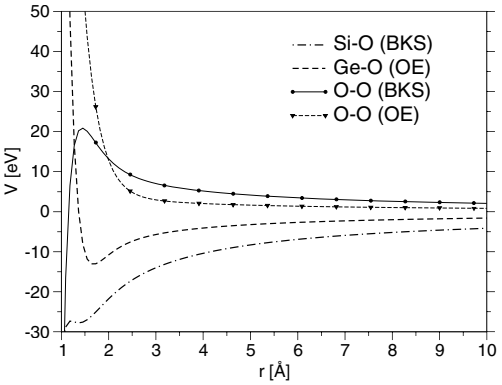


Fig. 1. Comparison of the Ge–O/Si–O and the O-O interaction potential as given by the BKS model for SiO_2 and the OE model for GeO_2.

3 Methods: A combination of classical and Car–Parrinello MD

In Car–Parrinello MD (CPMD) [CP85] the electronic degrees of freedom are explicitly taken into account via density functional theory. Thus, CPMD does not rely anymore on a empirical potential such as Eq. (1) but the potential between the atoms is computed "on the fly". The disadvantage of the Car–Parrinello technique is that it is about a factor 10^5 times slower than classical MD, and hence only very short time scales and very small systems are accessible. However, the method has been successfully applied in many applications in material science and quantum chemistry, and it has proven to be a very powerful tool in these areas (see references in [MH00]).

Therefore, we aim at parametrising a model potential for SiO_2–GeO_2 mixtures using the following strategy: CPMD is used to reparametrise the OE potential for GeO_2 such that the O–O potential is identical to the O–O potential in the BKS model and such that the partial charges of the Coulomb term coincide with those of the BKS potential. Thereby, it might be necessary to add a three–body term in order to stabilise the intra–tetrahedral O–Ge–O angle. For the Si–Ge interaction in the mixed system, we will then use Lorentz–Berthelot combination rules that are commonly known from Lennard–Jones models.

Once we have a potential model for the mixture, we can run a classical MD with this potential in order to check how accurate it is with respect to experimental data. This information can then be used again in the parametrisation procedure by means of CPMD. We hope that we will be able to develop an accurate potential of SiO_2–GeO_2 mixtures with such a combination of classical MD and CPMD.

One might tend to the close–at–hand idea to fit the potential of the SiO_2- GeO_2 mixture directly using CPMD. The main problem of this approach is the number of atoms to be involved. In order to get a good fit of the forces the sample has to be large enough to yield reasonable statistics (apart from finite size effects). E.g. a relatively large system size for a CPMD would be a system of 120 particles. Now, if

one considers a mixture with 25% SiO_2 and 75% GeO_2, there would be only *ten* Si atoms in the sample.

4 Results

The results presented in this section are devoted to a comparison of the structural and dynamic properties of amorphous SiO_2 and GeO_2. Whereas detailed simulation studies exist already for SiO_2 using the BKS potential (e.g. see [HK99, SSP01] and references therein), in the case of GeO_2 such simulations are still lacking. Therefore, we have done extensive MD simulations with the OE model for a system of $N = 1152$ particles at constant ambient pressure ($p = 0$). We equilibrated the system in the NPT ensemble followed by production runs in the microcanonical ensemble. Further details of the simulation can be found elsewhere [HHB].

Structure.

Fig. 2 shows the neutron scattering structure factor $S_{\mathrm{n}}(q)$ at $T = 300$ K (i.e. far below the glass transition temperature, see [HK99]) as obtained from the simulation in comparison to experimental data [SBL03, PC92]. The quantity $S_{\mathrm{n}}(q)$, depending on wavenumber q, can be estimated from the simulation by

$$S_{\mathrm{n}}(q) = \frac{N}{\sum_\alpha N_\alpha b_\alpha^2} \sum_{\alpha,\beta} b_\alpha b_\beta S_{\alpha\beta}(q) \qquad \alpha, \beta \in \{\mathrm{Si}, \mathrm{Ge}, \mathrm{O}\} \tag{2}$$

with the partial structure factors defined by

$$S_{\alpha\beta}(q) = \frac{1}{N} \left\langle \sum_{j=1}^{N_\alpha} \sum_{k=1}^{N_\beta} \exp(i\mathbf{q} \cdot \mathbf{r}_{jk}) \right\rangle \quad . \tag{3}$$

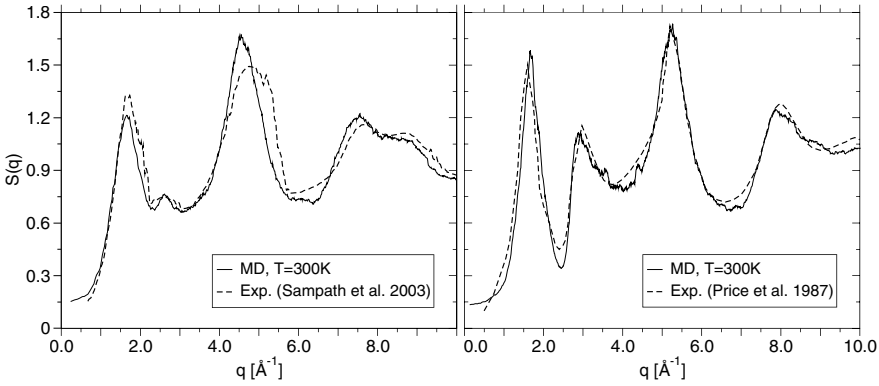


Fig. 2. Neutron scattering structure factor of GeO_2(left) and SiO_2(right): Simulation data in comparison to neutron scattering experiments.

In Eqs. (2) and (3), N_α is the number of particles of type α, and b_α denotes the scattering length of an particle of type α that we have taken from the literature.

As we can infer from Fig. 2 the simulation reproduces the experimental data very well both for SiO_2 and GeO_2 which indicates that the used pair potentials allow an accurate description of the latter systems in their amorphous state.

Fig. 3 shows the partial structure factors for the O–O correlation in SiO_2 and GeO_2 as obtained from the simulation. Especially interesting is the peak around $q = 1.6$ Å^{-1} which indicates the intermediate range order in both systems that stems from the tetrahedral network structure. As we see, this peak (as well as the second) peak is slightly shifted to lower values of q in GeO_2 which shows that the typical length scales of chemical ordering in GeO_2 are slightly larger than those in SiO_2. A detailed discussion of the structure of SiO_2 and GeO_2 can be found elsewhere [HHB].

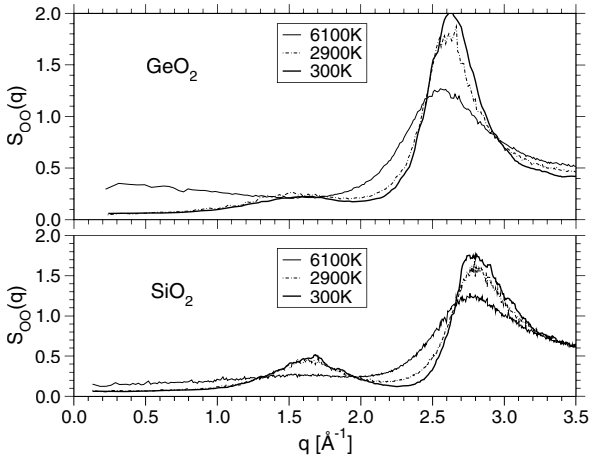


Fig. 3. Partial structure factor $S_{\mathrm{OO}}(q)$ at the indicated temperatures for GeO_2 (upper plot) and SiO_2 (lower plot).

Dynamics.

The self diffusion constants can be easily estimated in a MD simulation via the Einstein-relation $D_\alpha = \lim_{t\to\infty} \left\langle r_\alpha^2(t) \right\rangle / 6t$ where $\left\langle r_\alpha^2(t) \right\rangle$ is the mean–squared displacement for a tagged particle of type α. Fig. 4 shows the temperature dependence of the self diffusion constants for SiO_2 and GeO_2 in an Arrhenius plot. Obviously, below 4300K the dynamics of GeO_2 is increasingly faster than the one of silica at identical temperatures. Recently, we have shown for the case of SiO_2 [HK99] that self diffusion constants exhibit a fragile–to–strong transition where the fragile regime at high temperatures is well described by mode coupling theory. It will be interesting to see whether this holds also for GeO_2.

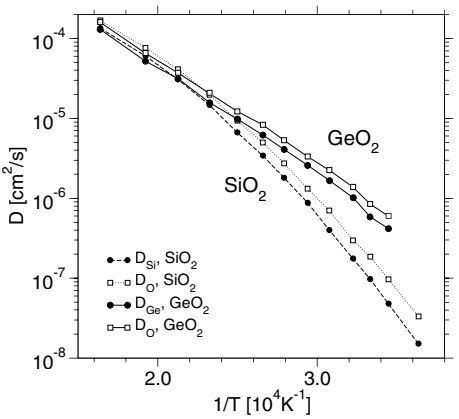


Fig. 4. Arrhenius plot of the diffusion constants of SiO_2 and GeO_2. The temperature ranges from 6100K to 2750K.

5 Summary

In this paper we have outlined a strategy to develop a potential function for SiO_2–GeO_2 mixtures by means of a combination of an *ab initio* technique (CPMD) and classical MD. Moreover we have presented results of a MD simulation of GeO_2 to highlight some of the subtle differences in the structural and dynamic properties of GeO_2 and SiO_2.

References

[BKS90] B. W. H. van Beest, G. J. Kramer, and R. A. van Santen: *Force fields for silicates and aluminophosphates based on ab initio calculations.* Phys. Rev. Lett. **64**, 1955 (1990).

[CP85] R. Car and M. Parrinello: *Unified Approach for Molecular dynamics and Density Functional Theory.* Phys Rev. Lett. **55**, 2471 (1985).

[HHB] M. Hawlitzky, J. Horbach, and K. Binder: *Structural and dynamic properties of germania: A molecular dynamics computer simulation.* In preparation.

[HK99] J. Horbach and W. Kob: *Static and dynamic properties of a viscous silica melt.* Phys. Rev. B **60**, 3169 (1999).

[HKB02] J. Horbach, W. Kob, and K. Binder: *Dynamics of sodium in sodium disilicate: Channel relaxation and sodium diffusion.* Phys. Rev. Lett. **88**, 125502 (2002).

[JKJ01] P. Jund, W. Kob, and R. Jullien, *Channel diffusion of sodium in a silicate glass.* Phys. Rev. B **64**, 134303 (2001).

[LKH03] H. Lammert, M. Kunow, and A. Heuer, *Complete identification of alkali sites in ion conducting lithium silicate glasses: A computer study of ion dynamics.* Phys. Rev. Lett. **90**, 215901 (2003).

[MH00] D. Marx and J. Hutter: *Ab initio moecular dynamics: Theory and Implementation.* Modern Methods and Algorithms of Quantum chemistry, NIC Series, Jülich, Vol. 1, 301-449 (2000).

[OE98] R. D. Oeffner and S. R. Elliott: *Interatomic potential for germanium dioxide empirically fitted to an ab initio energy surface.* Phys Rev. B **58**, 14791 (1998).

[PC92] D. L. Price and J. M. Carpenter: *Scattering function of vitreous silica.* J. Non-cryst. Sol. **92**, 153 (1987).

[SBL03] S. Sampath, C. J. Benmore, K. M. Lantzky, J. Neuefeind, K. Leinenweber, D. L. Price, J. L. Yarger: *Intermediate-Range Order in Permanently Densified* GeO_2*Glass.* Phys. Rev. Lett. **90**, 115502 (2003).

[SSP01] I. Saika–Voivod, F. Sciortino, and P. H. Poole: *Computer simulations of liquid silica: Equation of state and liquid–liquid phase transition.* Phys. Rev. E **63** 011202 (2001).

[WHK03] A. Winkler, J. Horbach, W. Kob, K. Binder: *Structure and Diffusion in an amorphous aluminium silicate: A Molecular Dynamics computer simulation.* J. Chem. Phys. (in press). Preprint: cond-mat/0305227.

Large Scale Density Functional Calculations

Jürg Hutter

Physical Chemistry Institute, University of Zürich,
Winterthurerstrasse 190, 8057 Zürich, Switzerland
hutter@pci.unizh.ch

Summary. New developments in algorithms for density functional calculations within the Kohn–Sham methods allow to study systems with several hundreds of atoms. We present a linear scaling method for the construction of the Kohn–Sham Hamiltonian based on fast Fourier transforms. To solve the Kohn–Sham equation the orbital rotation method provides an efficient scheme for small to medium sized systems, where methods depending on the sparsity of the density matrix are not yet applicable. Combining these methods with multiscale algorithms will make it possible to access length and time scales relevant for many problems in materials science, life sciences or catalysis.

1 Introduction

Ab initio molecular dynamics (AIMD) [CP85] is becoming a well established tool for the study of physical [CCS+99], chemical [Rot01] and biochemical [CR01] systems. This method has enabled direct simulations of chemical reactions in extended and complex environments. However, time and length scales accessible to *ab initio* molecular dynamics simulations are necessarily limited. This is the price one pays for an in principle unbiased description of chemical reactivity.

New developments in algorithms and more powerful computers make it possible to simulate larger systems for longer times. Nevertheless, there is the need for simulations of even larger systems as there is the need to be able to go beyond the time scales accessible by todays methods. The simulation of the electronic structure of biomolecules (especially DNA), systems with long range interactions (strongly ionic systems), electrochemical systems, or the behavior of nano structures are of major importance. Systems of this size will not be accessible with traditional algorithms used in electronic structure theory.

Algorithms that make use of the multiple time and length scales inherent in most physical systems are therefore needed to address many important large scale problems. However, at the center of these methods *ab initio* electronic structure calculations are most likely needed in order to achieve a high level of accuracy and transfer-

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

ability. It is therefore also in the context of multiscale modeling important to further develop *ab initio* methods.

The Kohn–Sham approach [KS65] to density functional [HK64, PY89] theory combines high computational efficiency and accuracy for electronic ground states. It is therefore the method of choice for many applications and is subject to intense research both on theoretical and algorithmic level. Current implementations of algorithms are optimized for small to medium sized systems (up to ca. 100 atoms). The standard Car–Parrinello programs [MH00, CPM] for example are based on a pseudopotential/plane wave approach that allows for efficient calculations of medium sized systems with periodic boundary conditions on massively parallel computers. However, the cubic scaling of these implementation prohibits their application to much larger systems.

Recently, much work has been done to devise new algorithms that reduce the scaling behavior and the ultimate goal, linear scaling has been achieved for Kohn–Sham density functional calculations and most other electronic structure methods. The review of Goedecker [Goe99] gives an overview of the different attempts.

2 Kohn–Sham Method

The Kohn–Sham energy is defined as

$$\mathrm{E_{KS}} = E_{\mathrm{kin}}[\{\Phi\}] + E_{\mathrm{ext}}[n] + E_{\mathrm{H}}[n] + E_{\mathrm{XC}}[n] \ , \tag{1}$$

where the kinetic energy E_{kin} is a function of the Kohn–Sham orbitals $\{\Phi\}$ and the external energy E_{ext}, Hartree energy E_{H}, and exchange and correlation energy E_{XC} are functions of the electron density n. The election density is calculated from the orbitals

$$n(\mathbf{r}) = \sum_{i=1}^{N} f_i |\Phi_i(\mathbf{r})|^2 \ , \tag{2}$$

and the occupation numbers f_i, where N is the number of occupied orbitals. Using the variational principle and imposing the orthogonality constraint on the orbitals leads to the Kohn–Sham equations (in canonical form)

$$H_{\mathrm{KS}}\Phi_i = \epsilon_i \Phi_i \ . \tag{3}$$

As the Kohn–Sham Hamiltonian H_{KS} depends on the orbitals, these equations have to be solved iteratively until self–consistence is achieved.

Expanding the orbitals in a basis set transforms the equation into algebraic form and we are left with the tasks to calculate the Kohn–Sham Hamiltonian in its matrix representation from a given set of orbitals and the search for the eigenvalues and eigenfunctions of this matrix. In the following sections we will discuss methods and algorithms to efficiently perform these two tasks.

To be able to calculate the Kohn–Sham matrix in a number of steps proportional to the system size, first it is necessary to find a matrix representation with only $O(N)$

non–vanishing elements. This can be achieved by using a basis set with finite support. There are two basis sets of this type used in linear scaling calculations. Basis sets connected with grids in real space include finite difference methods [CTS94], finite element type methods based on B-splines [HGG97] and plane wave derived basis sets [GHP01]. Another possibility is to use basis sets derived from atomic orbitals. These basis sets are used in quantum chemistry calculations and are very popular with linear scaling approaches [LHP97, OAS96].

Expanding atomic orbitals in linear combinations of Gaussian functions has the advantage that most of the integrals appearing in electronic structure calculations can be determined analytically. This together with the decay properties of Gaussians, both in real and reciprocal space make them an ideal choice for large scale density functional calculations.

2.1 Kohn–Sham Hamiltonian in a Gaussian Basis

Kohn–Sham orbitals are expanded in a set of basis functions

$$\Phi_i(\mathbf{r}) = \sum_{\alpha=1}^{M} c_{\alpha i}\varphi_\alpha(\mathbf{r}) \ , \tag{4}$$

where the M basis functions are centered at an atom and have the form of a polynomial multiplied by a Gaussian function. The Kohn–Sham Hamiltonian in this basis can be calculated from

$$H_{\mu\nu} = \langle\mu| -\frac{1}{2}\nabla^2|\nu\rangle + \langle\mu|V_{\rm ext}(\mathbf{r})|\nu\rangle + \langle\mu|V_{\rm xc}(\mathbf{r})|\nu\rangle + \frac{1}{2}\sum_{\alpha\beta} P_{\alpha\beta}\langle\mu\nu||\alpha\beta\rangle \ , \tag{5}$$

where $V_{\rm ext}$ is the external potential and $V_{\rm xc}$ the exchange and correlation potential. Coulomb integrals are defined by

$$\langle\mu\nu||\alpha\beta\rangle = \int\int \frac{\varphi_\mu(\mathbf{r})\varphi_\nu(\mathbf{r})\varphi_\alpha(\mathbf{r}')\varphi_\beta(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, d\mathbf{r}\, d\mathbf{r}' \ . \tag{6}$$

There are M^4 integrals of this type, but using screening techniques [HWH+91] the asymptotic number of non–vanishing integrals for large systems is M^2. To further reduce the scaling of this part of the calculation, adaptations of fast multipole methods to charge distributions have been developed. [WJGHG94]

Another approach expands the charge distribution into an auxiliary basis [SF75]

$$n(\mathbf{r}) = \sum_{\alpha\beta} P_{\alpha\beta}\varphi_\alpha(\mathbf{r})\varphi_\beta(\mathbf{r}) = \sum_k d_k\chi_k(\mathbf{r}) \ . \tag{7}$$

This reduces the inherent scaling for the Coulomb integrals to M^3, which can be again lowered to M^2 using integral screening. The advantage is that for small and medium system sizes, where the screening is not yet efficient, considerable speed up can be achieved compared to the original method. [ETO+95]

A plane wave expansion of the electronic density can be viewed as a special auxiliary basis set. Plane waves have the advantage of being orthogonal and the expansion coefficients do not depend on the metric used. Further do plane waves describe the Hartree potential derived from the density exactly in the same basis and the calculation of the potential only needs O(M) steps. The major disadvantage is the number of plane waves needed to accurately describe the density. This effectively limits its use to methods where the strongly varying parts of the electron density are either handled by pseudopotentials or taken care of in other ways. [LHP97, LHP99, KP00, FMP02a, FMP02b]

Another advantage of a plane wave expansion is that the long range part of the external potential can be treated together with the other Coulombic terms, greatly reducing the prefactor of the calculation of those matrix elements.

Accuracy and efficiency of the above described approach highly depends on the decay properties of Gaussian functions in real and reciprocal space. The fast decay of a Gaussian function in real space efficiently restricts the integration volume. The Fourier transform of a Gaussian function is again a Gaussian function and has therefore a fast decay in reciprocal space. This means that an accurate integration in real space can be achieved with a relatively small number of integration points (see figure 1).

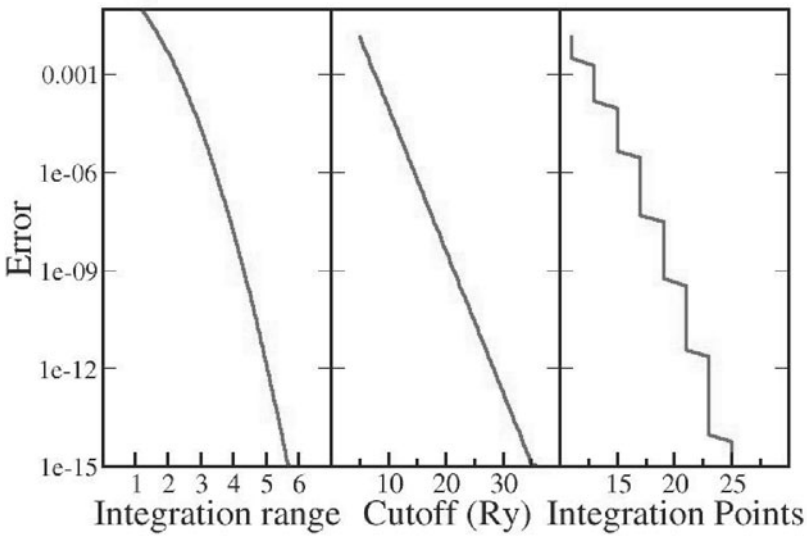


Fig. 1. Accuracy of real space integration as a function of integration radius [bohr], plane wave cutoff and number of integration points for a one-dimensional Gaussian function with an exponent of one.

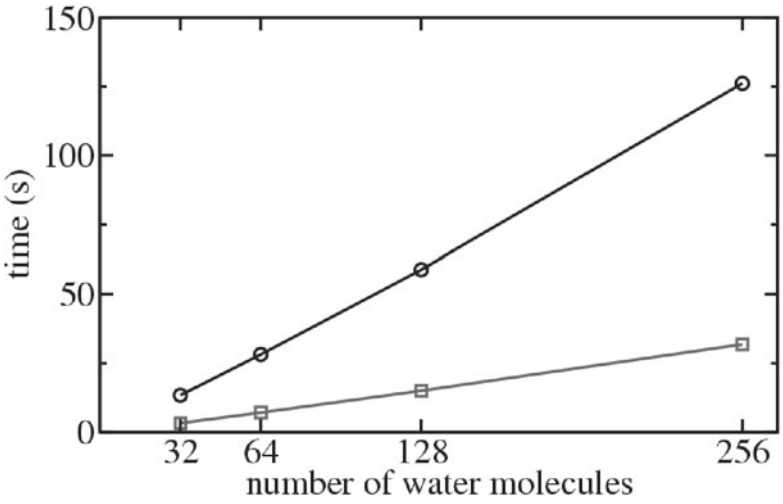


Fig. 2. Scaling of the construction of the Kohn–Sham Hamiltonian for periodic systems of water molecules. Circles: total time; squares: time spent in FFT routines.

The combination of efficient screening of integrals with the use of fast Fourier transformation to calculate the potentials leads to a scheme that shows near linear scaling [LHP97, LHP99, KP00]. In fact, it can be shown (see figure 2) that the onset of linear scaling is already reached for rather small systems. This is one of the main advantages of the FFT based method over the fast multipole methods.

2.2 Solving the Kohn–Sham Equations

Optimizing the Kohn–Sham energy with respect to the Kohn–Sham orbitals subject to the orthogonality constraint of the orbitals is traditionally done by diagonalizing the Kohn–Sham Hamiltonian together with a mixing/extrapolation scheme for the orbitals. These methods show a O(M^3) scaling with a small prefactor. They are highly efficient for small numbers of basis functions. However, when system sizes are increasing diagonalization time becomes rapidly the time limiting step.

The one-particle density operator (occupation numbers f_i)

$$P(\mathbf{r}, \mathbf{r}') = \sum_{i=1}^{N} f_i \Phi_i(\mathbf{r}) \Phi_i^*(\mathbf{r}') \tag{8}$$

has for insulators an exponential decay behavior for large values of $|\mathbf{r} - \mathbf{r}'|$. Expressing $P(\mathbf{r}, \mathbf{r}')$ in a local basis should therefore result in a sparse matrix representation. Using this fact, linear scaling minimization schemes to determine directly the density matrix P have been developed. However, these methods have a rather large prefactor that gets even larger for non-orthogonal atomic orbital based basis sets. This problem makes fully linear scaling applications using large basis sets impractical for system sizes accessible today.

One approach that addresses this problem is the method based on polarized atomic orbitals [LHG97, BPH02]. By introducing a variationally optimized intermediate minimal basis it was possible to almost achieve the accuracy of the underlying larger basis while reducing considerably the prefactor of the optimization method. Another route is to optimize cubically scaling algorithms further and so to extend the range of system sizes that can be treated. The orbital transformation method [VH03] was developed in this spirit.

We express N occupied orthonormal wave functions $\Phi_i(r)$ as a linear combination of M basis functions $\varphi_\alpha(r)$ using an $M \times N$ matrix $\mathbf{c}$ of orbital coefficients $c_{\alpha i}$. The calculation of the electronic structure involves the minimization of the electronic energy $E(\mathbf{c})$ with respect to $\mathbf{c}$ subject to the constraint that $\mathbf{c}^T\mathbf{S}\mathbf{c} = \mathbf{I}$. $\mathbf{S}$ is the overlap matrix ($S_{\alpha\beta} = \int \varphi_\alpha(r)\varphi_\beta(r)dr$) and $\mathbf{I}$ the identity matrix. New variables to describe $\mathbf{c}$, inspired by the form of the exponential transformations [HJO00] as given in Ref. [HPV94], are introduced. The new variables $\mathbf{x}$ are linearly constrained according to

$$\mathbf{x}^T\mathbf{S}\mathbf{c}_0 = 0 \ . \tag{9}$$

The constant initial vectors $\mathbf{c}_0$ fulfill $\mathbf{c}_0^T\mathbf{S}\mathbf{c}_0 = \mathbf{I}$. $\mathbf{c}$ is related to $\mathbf{x}$ by the following transformation

$$\mathbf{c}(\mathbf{x}) = \mathbf{c}_0 \cos(\mathbf{U}) + \mathbf{x}\mathbf{U}^{-1}\sin(\mathbf{U}) \ , \tag{10}$$

where the matrix $\mathbf{U}$ is defined as

$$\mathbf{U} = \left(\mathbf{x}^T\mathbf{S}\mathbf{x}\right)^{1/2} \ . \tag{11}$$

The functions of the symmetric matrix $\mathbf{x}^T\mathbf{S}\mathbf{x}$ are defined as functions of its eigenvalues. It can be verified that $\mathbf{c}^T(\mathbf{x})\ \mathbf{S}\ \mathbf{c}(\mathbf{x}) = \mathbf{I}$ for all choices of $\mathbf{x}$. The variables $\mathbf{x}$ can be used to optimize the energy $E(\mathbf{c}(\mathbf{x}))$ using any standard minimization algorithm. This is possible because $\mathbf{x}$ is linearly constrained which implies that the space of allowed $\mathbf{x}$ is a linear space. Therefore, a finite step along the gradient will produce a new point that still fulfills the constraint condition exactly. Hence, it is unnecessary to follow a curved geodesic during minimization, as is appropriate for variables that are constrained non-linearly such as $\mathbf{c}$.[VHG02, EAS98] The gradient of the energy with respect to the new variable $\mathbf{x}$ is given by (Lagrange multiplier Λ)

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}}\left[E(\mathbf{c}) + \mathrm{Tr}(\mathbf{x}^T\mathbf{S}\mathbf{c}_0\Lambda)\right] \\ = \frac{\partial E}{\partial \mathbf{c}}\frac{\partial \mathbf{c}}{\partial \mathbf{x}} - \mathbf{S}\ \mathbf{c}_0\ [(\mathbf{S}\ \mathbf{c}_0)^T\ \mathbf{S}\ \mathbf{c}_0]^{-1}(\mathbf{S}\ \mathbf{c}_0)^T\frac{\partial E}{\partial \mathbf{x}} \ .\end{aligned} \tag{12}$$

The gradient $\partial E/\partial \mathbf{x}$ is given by

$$\frac{\partial E}{\partial \mathbf{x}} = (\mathbf{H}\mathbf{c})\mathbf{U}^{-1}\sin(\mathbf{U}) + (\mathbf{S}\mathbf{x})\left(\mathbf{R}(\mathbf{K}^T + \mathbf{K})\mathbf{R}^T\right) \tag{13}$$

$$\mathbf{K} = \left(\mathbf{R}^T\left((\mathbf{H}\mathbf{c}^T\mathbf{x})\,\mathbf{R}\right) \otimes \mathbf{D}^1 + \left(\mathbf{R}^T\left((\mathbf{H}\mathbf{c})^T\mathbf{c}_0\right)\mathbf{R}\right) \otimes \mathbf{D}^2 \tag{14}$$

The notation $\otimes$ is used for the direct matrix product $(\mathbf{A} \otimes \mathbf{B})_{ij} = A_{ij}B_{ij}$, $\mathbf{HC}$ represents $\partial E/\partial \mathbf{c}$, $\mathbf{R}$ is the matrix of eigenvectors of $\mathbf{x}^T\ \mathbf{S}\ \mathbf{x}$, and $\mathbf{D}^1$ and $\mathbf{D}^2$ are given by

$$D^1_{ij} = \frac{\frac{\sin(\sqrt{\Lambda_i})}{\sqrt{\Lambda_i}} - \frac{\sin(\sqrt{\Lambda_j})}{\sqrt{\Lambda_j}}}{\Lambda_i - \Lambda_j} \qquad D^2_{ij} = \frac{\cos(\sqrt{\Lambda_i}) - \cos(\sqrt{\Lambda_j})}{\Lambda_i - \Lambda_j} \tag{15}$$

where Λ_i is the eigenvalue corresponding to the i^{th} column of $\mathbf{R}$. The calculation of the gradient involves the computation of sparse matrix - full matrix products in order to compute $\mathbf{Hc}$ and $\mathbf{Sx}$, and several full matrix - full matrix products such as $(\mathbf{Hc})^T\mathbf{x}$ or $\mathbf{x}f(\mathbf{U})$. For large systems, they have a computational cost of $O(MN)$ and $O(MN^2)$ respectively. Notice that the full matrix products run almost with peak performance on modern computers, and that they are well parallelized. In case the matrices $\mathbf{H}$ and $\mathbf{S}$ are not yet sparse the cost of computing $\mathbf{Hc}$ and $\mathbf{Sx}$ is $O(M^2N)$. The diagonalization of the $N \times N$ matrix $\mathbf{x}^T\ \mathbf{S}\ \mathbf{x}$, which is needed to compute the matrix functions efficiently, costs only a small fraction of the total CPU time.

For practical applications it is important to precondition the minimization. An effective preconditioner can be obtained by approximating the electronic Hessian

$$\frac{\partial^2 E}{\partial x_{\alpha i}\partial x_{\beta j}} \approx 2H_{\alpha\beta}\delta_{ij} - 2S_{\alpha\beta}\delta_{ij}\epsilon^0_i\ , \tag{16}$$

where we assume that $(\mathbf{c}_0^T\ \mathbf{H}\ \mathbf{c}_0)_{ij} = \epsilon^0_i\delta_{ij}$. This preconditioner is symmetric and positive definite for all $\mathbf{x}$ that satisfy $(\mathbf{x}^T\mathbf{S}\mathbf{c}_0 = 0)$) if $\mathbf{c}_0$ is close to the minimum.

It would be optimal to use a preconditioned gradient where every gradient vector is multiplied with a different matrix (dependence on ϵ_i). In practice, a single symmetric positive definite matrix $\mathbf{P}$ is constructed so that $\mathbf{P}(\mathbf{H} - \mathbf{S}\epsilon)\mathbf{x} - \mathbf{x} \approx 0$. Since introducing a preconditioner is equivalent to a change in variables $\mathbf{x} \rightarrow \sqrt{\mathbf{P}}\mathbf{x}$ the Lagrangian multipliers have to be adapted accordingly.

3 Examples

To test the methods described in the previous sections calculations on water systems were performed. In table 1 the time to calculate the forces on all atoms for periodic systems of water molecules are listed. From the scaling between two sizes of systems we can see that the orbital transformation method has a very small prefactor for the cubic part and scales for systems up to 1500 atoms essentially quadratic.

The second system, which is based on a X-ray structure of synthetic DNA, has been studied before by Gervasio et. al (Ref. [GCP02]). We have doubled the system so that two times twelve base-pairs, solvent and counter ions have been described, a total of 2388 atoms. The system has a total of 3960 occupied orbitals and DZV(d,p) and TZV(2d,2p) basis sets (22596 and 38688 basis functions) have been tested. A density mapping accuracy of 10^{-12} has been chosen and elements of the overlap

Table 1. CPU time for the calculation of forces on all atoms of different water systems with periodic boundary conditions. All times in seconds on 16 CPU IBM p690.

H_2O	basis	states	time	scaling
32	736	128	46	–
64	1472	256	108	1.23
128	2944	512	281	1.37
256	5888	1024	854	1.60
512	11776	2048	3294	1.95
1024	23552	4096	15833	2.26

matrix have been neglected if they were estimated to be less than 10^{-10}. The sparsity pattern of the overlap matrix has also been used for the Kohn-Sham matrix. The wavefunctions have been optimized using the orbital transformation method with a conjugate gradient minimizer and a kinetic energy based preconditioner. A minimal basis set has been used to initialize the wavefunction. The calculations have been performed on 32 CPUs of an IBM p690 (Power4 CPUs, 1.3GHz), and needed about 2.5 and 5 hours to converge. The time needed for a single two point line search was 675 and 1100 seconds for DZV(d,p) and TZV(2d,2p) respectively. These timings reflect accurately the effective linear scaling in the number of basis functions for constant system size. 14 line searches were necessary to reach convergence (gradients $2 \cdot 10^{-5}$). Interestingly, even for a system of this size, the cubically scaling part is not yet fully dominant. For the TZV(2d,2p) system, 45, 43, and 8 percent of the total time is spent in the cubic, the quadratic and the linear part respectively. This is in part an effect of the superior parallel efficiency of the cubic part. The timing for the cubic part includes the time needed for constructing and applying the preconditioner, 10 and 15 percent of the total time respectively. Another observation of relevance is related to the sparsity of the density matrix. We find that over 60 percent of the atomic blocks of the density matrix are non-zero if a tolerance of only 10^{-5} is used. Even if such a moderate accuracy is demanded, linear scaling methods that rely on the sparsity of the density matrix cannot be efficient.

Acknowledgement. I would like to thank Joost VandeVondele and Matthias Krack for providing the example calculations.

References

[BPH02] G. Berghold, M. Parrinello, and J. Hutter. Polarized atomic orbitals for linear scaling methods. J. Chem. Phys., 116:1800, 1810, 2002.

[CCS+99] C. Cavazzoni, G. L. Chiarotti, S. Scandalo, E. Tosatti, M. Bernasconi, and M. Parrinello. Superionic and metallic states of water and ammonia at giant planet conditions. Science, 283:44, 46, 1999.

[CP85] R. Car and M. Parrinello. Unified Approach for Molecular Dynamics and Density-Functional Theory. Physical Review Letters, 55:2471, 2474, 1985.

[CPM] CPMD V3.7 Copyright IBM Corp 1990-2003, Copyright MPI fur Festkorperforschung Stuttgart 1997-2001. see also www.cmpd.org.

[CR01] P. Carloni and U. Rothlisberger. Simulations of enzymatic systems: Perspectives from Car-Parrinello molecular dynamics simulations. In L. Eriksson, editor, Theoetical Biochemistry - Processes and Properties of Biological Systems, page 215. Elsevier Science, 2001.

[CTS94] J. R. Chelikowsky, N. Troullier, and Y. Saad. Finite-differencepseudopotential method - electronic-structure calculations without a basis. Phys. Rev. Lett., 72:1240, 1243, 1994.

[EAS98] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl., 20:303, 353, 1998.

[ETO+95] K. Eichkorn, O. Treutler, H. Ohm, M. Haser, and R. Ahlrichs. Auxiliary basis-sets to approximate coulomb potentials. Chem. Phys. Lett., 240:283, 289, 1995.

[FMP02a] L. Fusti-Molnar and P. Pulay. Accurate molecular integrals and energies using combined plane wave and Gaussian basis sets in molecular electronic structure theory. J. Chem. Phys., 116:7795, 7805, 2002.

[FMP02b] L. Fusti-Molnar and P. Pulay. The fourier transform coulomb method: E.cient and accurate calculation of the coulomb operator in a Gaussian basis. J. Chem. Phys., 117:7827, 7835, 2002.

[GCP02] F. L. Gervasio, P. Carloni, and M. Parrinello. Electronic structure of wet DNA. Phys. Rev. Lett., 89:108102, 2002.

[GHP01] C. K. Gan, P. D. Haynes, and M. C. Payne. First-principles densityfunctional calculations using localized spherical-wave basis sets. Phys. Rev. B, 63:205109, 2001.

[Goe99] S. Goedecker. Linear scaling electronic structure methods. Rev. Mod. Phys., 71:1085, 1123, 1999.

[HGG97] E. Hern.andez, M. J. Gillan, and C. M. Goringe. Basis functions for linear-scaling first-principles calculations. Phys. Rev. B, 55:13485, 13493, 1997.

[HJO00] T. Helgaker, P. J.rgensen, and J. Olsen. Molecular Electronic-structure Theory. John Wiley & Sons Ltd, Chichester, 2000.

[HK64] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. Phys. Rev., 136:B864, B871, 1964.

[HPV94] J. Hutter, M. Parrinello, and S. Vogel. Exponential transformation of molecular-orbitals. J. Chem. Phys., 101:3862, 3865, 1994.

[HWH+91] H. Horn, H. Weiss, M. Haser, M. Ehrig, and R. Ahlrichs. Prescreening of 2-electron integral derivatives in SCF gradient and hessian calculations. J. Comp. Chem., 12:1058, 1064, 1991.

[KP00] M. Krack and M. Parrinello. All-electron ab-initio molecular dynamics. Phys. Chem. Chem. Phys., 2:2105, 2112, 2000.

[KS65] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. Phys. Rev., 140:A1133, A1139, 1965.

[LHG97] M. S. Lee and M. Head-Gordon. Polarized atomic orbitals for selfconsistent field electronic structure calculations. J. Chem. Phys., 107:9085, 9095, 1997.

[LHP97] G. Lippert, J. Hutter, and M. Parrinello. A hybrid Gaussian and plane wave density functional scheme. Mol. Phys., 92:477, 487, 1997.

[LHP99] G. Lippert, J. Hutter, and M. Parrinello. The Gaussian and augmentedplane- wave density functional method for ab initio molecular dynamics simulations. Theor. Chem. Acc., 103:124, 140, 1999.

[MH00] D. Marx and J. Hutter. ab-initio Molecular Dynamics: Theory and Implementation. In J. Grotendorst, editor, Modern Meth- ods and Algorithms of Quantum Chemistry, volume 1 of NIC Se- ries, pages 329, 477. FZ Julich, Germany, 2000. see also www.fz-juelich.de/nic-series/Volume1.

[OAS96] P. Ordej.on, E. Artacho, and J. M. Soler. Self-consistent order-N density-functional calculations for very large systems. Phys. Rev. B, 53:R10441, R10444, 1996.

[PY89] R. G. Parr and W. Yang. Density-Functional Theory of Atoms and Molecules. Oxford University Press, New York, 1989.

[Rot01] U. Rothlisberger. 15 years of Car-Parrinello simulations in physics,chemistry and biology. In J. Leszczynski, editor, Computational Chemistry: Reviews of Current Trends, pages 33, 68. World Scientific, Singapore, 2001.

[SF75] H. Sambe and R. H. Felton. New computational approach to slaters SCF-$\chi_a$equation. J. Chem. Phys., 62:1122, 1126, 1975.

[VH03] J. VandeVondele and J. Hutter. An e.cient orbital transformation method for electronic structure calculations. J. Chem. Phys., 118:4365 , 4369, 2003.

[VHG02] T. Van Voorhis and M. Head-Gordon. A geometric approach to direct minimization. Mol. Phys., 100:1713, 1721, 2002.

[WJGHG94] C. A. White, B. G. Johnson, P. M. W. Gill, and M. Head-Gordon. The continuous fast multipole method. Chem. Phys. Lett., 230:8, 16, 1994.

Dispersion corrected density functionals applied to the water naphthalene cluster

Urs Zimmerli[1], Michele Parrinello[2], and Petros Koumoutsakos[13]

[1] Institute of Computational Science, Department of Computer Science, ETH Zürich, 8092 Zürich, Switzerland
[2] Chair of Computational Science, Department of Chemistry, ETH Zürich
[3] Computational Laboratory, Department of Computer Science, ETH Zürich
petros@inf.ethz.ch

Summary. We study a recently proposed set of methods that extend Density Functional Theory (DFT) to include long-range dispersive interactions. The water naphthalene interaction is discussed and is compared to related results for water benzene clusters. The comparison is aimed at establishing an optimal representation for the water aromatic interaction within DFT that can be used to study the water carbon nanotube interaction.

1 Introduction

Density functional theory (DFT) has become a well-established method for electronic structure calculations of atoms, molecules and solids [Jon89] since it allows for the calculation of properties from first principles at moderate computational cost. While electrostatic interactions are well represented and even hydrogen bonds can be calculated with reasonable accuracy [Sim92], there are shortcomings of this method concerning the representation of long-range dispersion interactions [Kri94, Per95]. In particular, current density functionals fail to reproduce the leading r^{-6} dispersion contribution when density overlap becomes negligible [Per99a].

This failure can be attributed to their functional form that is based on local density, its gradient [Per92] and on the local kinetic energy density [Perd99], only. As dispersion forces become important where overlap of electron density is negligible, the density functionals fail to reproduce them by construction.

Several methods have been developed to overcome this problem. The approaches include a double local density approximation [Hul96, Hul99, And96, Rap91] which is, however, limited to non overlapping densities. Alternatively, Kohn *et al.* [Koh98] present an approach that is valid at all distances but which is also computationally very demanding. Another class of approaches is based on the introduction of a damped correction term [Wu01, Wu02] and has been applied to Hartree-Fock calculations [Dou82, Tan84] and to a density functional based, tight binding method

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

[Els01]. Recently, we reviewed this class of methods and compared their predictions on the interaction energy profile for the water benzene cluster [Zim03].

We apply the correction schemes reviewed in [Zim03] to DFT results of the water naphthalene cluster and verify the validity of the extrapolation to higher aromatics and eventually to graphite-like systems and carbon nanotubes. The interaction of graphite and carbon nanotubes with water are of particular interest in the design of nano devices and actuators and its magnitude is still unclear [WWJ+03].

This article is structured as follows. We first briefly review the correction schemes as presented by Elstner *et al.* [Els01], Wu *et al.* [Wu01] and Wu and Yang [Wu02], Followed by a description of the computational setup used for the water naphthalene cluster and a detailed discussion of the results. We conclude with the recommendation of a set of schemes whose performance compare well with data obtained from MP2 calculations.

2 Correction schemes

To represent the long-range dispersion interaction, we partition the total interaction energy

$$E_{tot} = E_{DFT} + \sum_{n} \sum_{\alpha>\beta} f_d(r_{\alpha\beta}) \frac{C_{\alpha\beta}}{r_{\alpha\beta}^6}. \tag{1}$$

Here E_{DFT} is the DFT interaction energy computed with an approximated exchange correlation functional and the second term is the dispersion energy represented as a two-body interaction. The indexes α and β are the centers of a pair of interacting particles and $r_{\alpha\beta}$ is the distance between them. $C_{\alpha\beta}$ is the interaction coefficient and f_d the related damping function. The particles considered for the dispersion energy contribution may either be molecules or atoms. This has an influence on the functional form of the contribution.

2.1 Molecular correction

A molecular representation of the dispersion energy contribution has been suggested by Wu *et al.* [Wu01], with a dispersion coefficient $C_{\alpha\beta}$ which is calculated as follows:

$$C_{\alpha\beta} = \frac{3}{2} \bar{p}_\alpha \bar{p}_\beta \frac{I_\alpha I_\beta}{I_\alpha + I_\beta}. \tag{2}$$

where $\bar{p}_\alpha$ denotes the average molecular polarizability of the molecule α and I_α the related ionization energy. To ensure physical behavior in the short range, the following damping function is suggested [Dou82]:

$$f_d(r) = \left[-\exp\left(-\frac{2.1r}{6} - 0.109 \frac{r^2}{6^{\frac{1}{2}}} \right) \right]^6 \tag{3}$$

2.2 Atomic correction

Atomic correction schemes have been proposed by Wu and Yang [Wu02] and Elstner *et al.* [Els01]. Both suggest the introduction of a dispersion energy term for every pair of atoms in the system. Atomic dispersion energy coefficients are calculated using the expression

$$C_{\alpha\alpha} = 0.75\sqrt{N_\alpha p_\alpha^3}. \tag{4}$$

Halgren [Hal92] proposes the following relationship to determine the Slater-Kirkwood effective number of electrons, N_α:

$$N_\alpha = 1.17 + 0.33 n_{\nu,\alpha}. \tag{5}$$

where, $n_{\nu,\alpha}$ is the number of valence electrons of atom α and N_α for hydrogen is set to 0.8. For diatomic coefficients $C_{\alpha\beta}$, Halgren [Hal92] recommends the Slater-Kirkwood combination rule

$$C_{\alpha\beta} = \frac{2C_{\alpha\alpha}C_{\beta\beta}p_\alpha p_\beta}{p_\alpha^2 C_{\beta\beta} + p_\beta^2 C_{\alpha\alpha}}. \tag{6}$$

Here p_α is the atomic polarizability of atom α. Elstner *et al.* [Els01] used equations (4-6) to compute their dispersion coefficients, while Wu and Yang [Wu02] propose a modified Slater-Kirkwood combination rule instead:

$$C_{6,\alpha\beta} = \frac{2\left(C_{6,\alpha\alpha}^2 C_{6,\beta\beta}^2 N_\alpha N_\beta\right)^{\frac{1}{3}}}{\left(N_\beta^2 C_{6,\alpha\alpha}\right)^{\frac{1}{3}} + \left(N_\alpha^2 C_{6,\beta\beta}\right)^{\frac{1}{3}}}. \tag{7}$$

Furthermore, Wu and Yang [Wu02] use atomic dispersion coefficients fitted to given molecular coefficients and propose the following two damping functions:

$$f_d(r) = \left(1 - \exp\left[-3.54\left(\frac{r}{R_m}\right)^3\right]\right)^2 \tag{8}$$

and

$$f_d(r) = \frac{1}{1 + \exp\left[-23\left(\frac{r}{R_m} - 1\right)\right]}. \tag{9}$$

Here, R_m is the sum of the atomic van der Waals radii obtained from Bondi [Bon64]. Elstner *et al.* [Els01] suggest the following damping function

$$f_d(r) = \left[1 - \exp\left(-3.0\left[\frac{r}{R_0}\right]^7\right)\right]^4, \tag{10}$$

where $R_0 = 3.8\,\text{Å}$ for first row elements.

3 Results

We evaluate the performance of one molecular and three atomic correction schemes. The density functionals considered in this work are B3LYP [Bec93, Lee88], PW91 [Per91], PBE [Per96] BLYP [Bec88, Lee88] HCTH/120 [Boe00]

For each scheme we compute interaction energy profiles for the water naphthalene cluster in eight different conformations. Two locations of the water molecule are considered, one with the oxygen atom centered on top of an aromatic ring and one with the oxygene atom centered on top of the carbon bond separating the two aromatic rings. In both of these locations two orientations of the water molecule are considered, the plane spanned by the water molecule is either orthogonal or in line with the main axis of the naphthalene molecule. All geometries, were evaluated, once with the water dipole pointing perpendicularly towards the naphthalene plane and once with the inverse orientation.

We do not consider geometries with the water dipole parallel to the aromatic plane as they do not provide additional information [Zim03]. Where not stated otherwise, all distances refer to the distance of the water oxygen to the naphthalene plane.

The performance of the correction schemes is quantified in two ways. On one hand a piecewise constant integration of the absolute deviation from the reference energy profile is performed to obtain the average error. On the other, during integration the absolute deviation is weighted with Boltzmann factors, $e^{-\beta V(r)}$, based on the reference potential $V(r)$ at the respective distance. We take $\beta = \frac{1}{k_B T}$ and set the temperature $T = 300\,\mathrm{K}$. The Boltzmann constant is represented by k_B.

DFT calculations are performed using CPMD [CPMD01], a plane wave DFT code. The computational cell was set to $25 \times 25 \times 50\, a_0$ for the calculations of DFT-based interaction energies. A cutoff of 90 Rydberg was applied for the plane wave basis and Trouiller-Martins pseudopotentials [Tro91] were used. The non-local part of the pseudopotentials are calculated according to the Kleinmann-Bylander scheme [Kle82].

The reference interaction energy profile is obtained by counterpoise-corrected MP2 calculations. The calculations were performed using Gaussian 98 [Gau98] with 6-311G(2d,2p) basis sets.

3.1 Uncorrected density functionals

The density functionals PW91, PBE and HCTH show reasonable agreement with the reference data for the geometries in which the dipole is pointing away from the naphthalene plane. However, for the opposite orientation of the water molecule, all functionals fail to reproduce the strength of the interaction reliably. The PW91 density functional results in the lowest mean deviation ($1.41\,\mathrm{kJ\,mol^{-1}}$) and the lowest Boltzmann weighted error of all uncorrected density functionals. Nevertheless, this functional underestimates the binding strength significantly for the geometries of low interaction energy, i.e. the geometries where the water dipole is pointing towards the

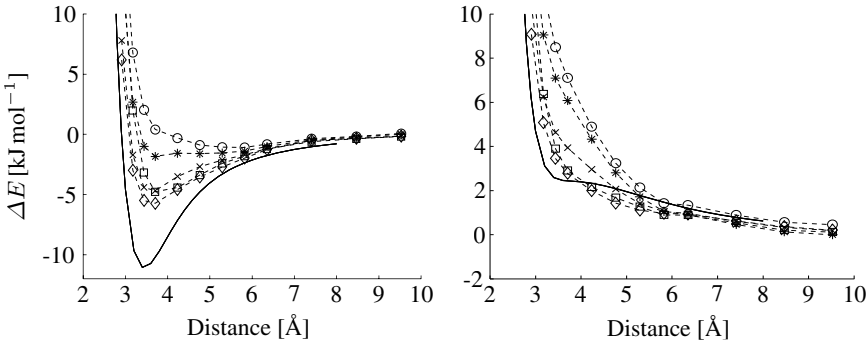


Fig. 1. The interaction energy profile for the water molecule centered on top of the bond separating the two aromatic rings of naphthalene with the dipole pointing perpendicularly towards (left) and away from (right) the naphthalene plane. The water plane is orthogonal to the main axis of the naphthalene. Some of the uncorrected density functionals show good agreement for the latter case whereas the attraction is clearly underestimated in the former. Solid line: reference data, dashed lines with symbols: DFT results, circles: BLYP, stars: B3LYP, squares: HCTH, crosses: PBE, diamonds: PW91.

naphthalene plane, (Figure 1). These results are fully consistent with the results reported for the water benzene cluster [Zim03].

3.2 Molecular Correction

The molecular correction does not account for anisotropies in the dispersion contribution due to the geometry of the molecule. It is therefore expected to break down at some critical molecular size.

For the water naphthalene interaction, we calculate a molecular dispersion coefficient of 17736 kJ mol^{-1} Å^6 using equation (2). The average polarizabilities are 1.5 Å^3 [IAPWS01] for water and 16.5 Å^3 [Hin93] for naphthalene. The ionization energies are 12.62 eV [Ton91] and 8.15 eV [Sch01], respectively.

This scheme applied to the HCTH functional leads to excellent results in the water benzene case: the lowest Boltzmann weighted error for all schemes is obtained. However, for the resulting interaction energy profiles significant deviations become apparent, especially for the geometries in which the water molecule is on top of the carbon carbon bond separating the two aromatic rings with the water dipole pointing away from the bond. The interaction, although it should be purely repulsive, becomes attractive for the HCTH functional (Figure 2). This would have a significant influence on the qualitative behavior of the system in molecular dynamics simulations, as the probability distribution for the relative position of a water molecule in the vicinity of naphthalene would be distorted.

In general, the dispersion contribution is too large for geometries in which the water molecule is centered on top of the bond which is connecting the two aromatic rings. This effect can be attributed to the functional form of the correction, which positions the two centers of interaction at the centers of mass of the molecules. While

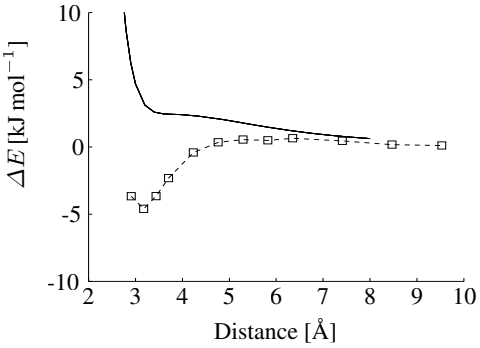


Fig. 2. The interaction energy profile for the water molecule centered on top of the carbon carbon bond separating the two aromatic rings of naphthalene with the dipole pointing away from the naphthalene plane. The plane spanned by the water molecule is orthogonal to the main axis of the naphthalene. The molecularly corrected HCTH functional (dashed line with squares) shows strong attraction with a deep minimum at approximately 3 Å, whereas the reference potential is purely repulsive (solid line).

geometries centered on top of the bond and on top of an aromatic ring result in very similar interaction energies on a MP2 level of theory, there are large discrepancies between the molecularly corrected DFT interaction energies. Both these effects can clearly be attributed to the breakdown of the molecular correction scheme that ignores the geometry of the molecules and their orientation with regard to each other.

3.3 Atomic Correction

In Zimmerli *et al.* [Zim03], for the water benzene system, it is recommended to use a corrected B3LYP density functional according to Wu and Yang [Wu02] along with a modified Slater-Kirkwood mixing rule (7) and damping function (9). For the water naphthalene cluster the agreement of the reference data and the corrected B3LYP functional is still good, with an average error of 1.25 kJ mol^{-1}. The Boltzmann weighted errors also indicate good agreement in the geometries with low interaction energy.

However, on the water naphthalene cluster, the correction scheme presented by Elstner *et al.* [Els01] using equations (6) and (10) applied to the PBE density functional, give the most consistent results with respect to the reference data. The average error is 1.04 kJ mol^{-1} and the Boltzmann weighted error is also one of the lowest in this study. Comparable results are also obtained if the PBE functional is corrected according to Wu and Yang [Wu02] using equations (7) and (9). While in the first case the dispersion contribution seems to be underaccounted for, there seems to be a tendency for overaccounting it in the latter case (Figure 3).

If the BLYP functional should be corrected for dispersion, the results of the current study indicate that the approach of Wu and Yang [Wu02] using equations (7) and (8) should be used. However, we also confirm that in this framework, the interaction energy is overestimated, resulting in an attraction which is roughly 3 kJ mol^{-1} too

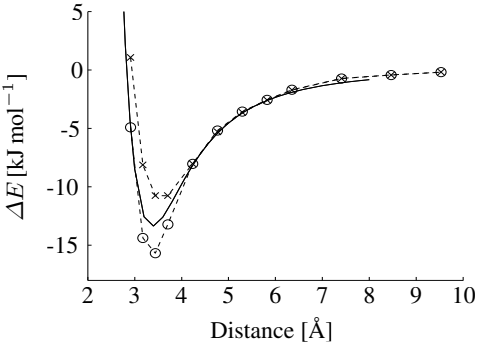


Fig. 3. The interaction energy profile for the water molecule centered on top of the bond joining the two aromatic rings of the naphthalene molecule with the dipole pointing towards the naphthalene plane. The plane spanned by the water molecule is parallel to the main axis of the naphthalene molecule. The solid line is the MP2-reference data, the dashed lines represent the corrected PBE functional, crosses indicate a correction according to [Els01], circles according to [Wu02] using damping function (9).

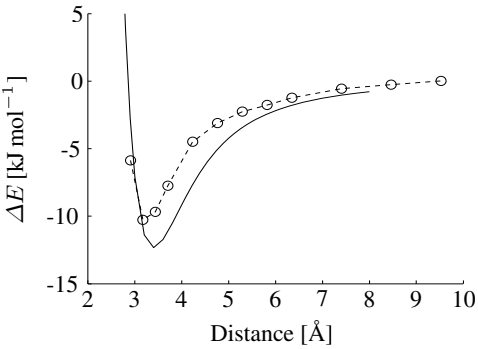


Fig. 4. The interaction energy profile for the water molecule centered on top of an aromatic ring of naphthalene with the dipole pointing towards the naphthalene plane. The plane spanned by the water molecule is parallel to the main axis of the naphthalene. The reference data (MP2) is represented by the solid line, the dashed line with circles represents the data obtained with the BLYP functional corrected according to [Wu02] using damping function (8). The interaction energy is overestimated in the range of 3.5 Å to 5.5 Å the absolute deviation however never exceeds 3 kJ mol^{-1}.

weak in the range of 3.5 Å to 5.5 Å, when the water dipole points towards the surface (see figure 4).

Except for functional BLYP, which has been discussed before, the correction according Wu and Yang [Wu02] using equations (7) and (9) gives better results than the ones damped with function (8).

4 Conclusion

The results of this paper indicate that accurate accounting of the dispersion interaction is crucial for the description of the interaction of water with aromatics. Current density functionals can give reasonable results for specific geometries. In particular the PW91 density functional results in a good approximation for geometries where the water dipole is pointing away from the naphthalene surface. However, as the quality of these results is geometry dependent, its predictive power is limited.

For the water naphthalene interaction, the molecular correction scheme presented in Wu *et al.* [Wu01] breaks down gradually, though still delivering good results in terms of average errors. Correction schemes based on atomic dispersion contributions are shown to be successful in predicting the interaction energy for the water naphthalene cluster in a consistent way. In particular the correction scheme suggested by Elstner *et al.* [Els01] performs well in combination with the PBE functional. Furthermore, the scheme by Wu and Yang [Wu02] performs well in combination with the B3LYP functional. For the calculation of the water aromatic interaction we therefore recommend the reader to use either of the latter correction schemes.

In general, the results reported by Zimmerli *et al.* [Zim03] are found to hold not only for the interaction of water with benzene but also for higher aromatics.

References

[Jon89] R. O. Jones and O. Gunnarsson. The density functional formalism, its application and prospects. *Rev. Modern Phys.*, 61(3):689–746, 1989.

[Sim92] Fiona Sim, Alain St-Amant, Imre Papai, and Dennis R. Salahub. Gaussian density functional calculations on hydrogen-bonded systems. *J. Am. Chem. Soc.*, 114(11):4391–4400, 1992.

[Kri94] Sándor Kristyán and Péter Pulay. Can (semi)local density functional theory account for the London dispersion forces? *Chem. Phys. Lett.*, 229:175–180, 1994.

[Per95] José Pérez-Jordá and A. D. Becke. A density-functional study of van der Waals forces: rare gas diatomics. *Chem. Phys. Lett.*, 233:134–137, 1995.

[Per99a] José Pérez-Jordá, Emilio San-Fabiaán, and Angel Pérez-Jiménez. Density-functional study of van der Waals forces on rare-gas diatomics: Hartree-Fock exchange. *J. Chem. Phys.*, 110(4):1916–1920, 1999.

[Per92] John P. Perdew, J. A. Chevary, S. H. Vosko, Koblar A. Jackson, Mark R. Pederson, D. J. Singh, and Carlos Fiolhais. Atoms, molecules, solids and surfaces: Application of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 46(11):6671–6687, 1992.

[Perd99] John P. Perdew, Stefan Kurth, Ales Zupan, and Peter Blaha. Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation. *Phys. Rev. Lett.*, 82(12):2544–2547, 1999.

[Hul96] Erika Hult, Ylva Andersson, Bengt I. Lundqvist, and David C. Langreth. Density functional for van der Waals forces at surfaces. *Phys. Rev. Lett.*, 77(10):2029–2032, 1996.

[Hul99] Erika Hult, Henrik Rydberg, Bengt I. Lundqvist, and David C. Langreth. Unified treatment of asymptotic van der Waals forces. *Phys. Rev. B*, 59(7):4708–4713, 1999.

[And96] Ylva Andersson, D. C. Langreth, and B. I. Lundqvist. van der Waals Interactions in Density-Functional Teory. *Phys. Rev. Lett.*, 76(1):102–105, 1996.

[Rap91] K. Rapcewicz and N. W. Ashcroft. Fluctuation attraction in condensed matter: A nonlocal functional approach. *Phys. Rev. B*, 44(8):4032–4035, 1991.

[Koh98] Walter Kohn, Yigal Meir, and Dmitrii E. Makarov. van der Waals energies in density functional theory. *Phys. Rev. Lett.*, 80(19):4153–4156, 1998.

[Wu01] X. Wu, M. C. Vargas, S. Nayak, V. Lotrich, and G. Scoles. Towards extending the applicability of density functional theory to weakly bound systems. *J. Chem. Phys.*, 115(19):8748–8757, 2001.

[Wu02] Qin Wu and Weitao Yang. Empirical correction to density functional theory for van der Waals interactions. *J. Comput. Phys.*, 116(2):515–524, 2002.

[Dou82] C. Douketis, G. Scoles, S. Marchetti, M. Zen, and A. J. Thakkar. Intermolecular forces via hybrid Hartree-Fock-SCF plus damped dispersion (HFD) energy calculations. An improved spherical model. *J. Chem. Phys.*, 76(6):3057–3063, 1982.

[Tan84] K. T. Tang and J. P. Toennies. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *J. Comput. Phys.*, 80(8):3726–3741, 1984.

[Els01] Marcus Elstner, Pavel Hobza, Thomas Frauenheim, Sándor Suhai, and Efthimios Kaxiras. Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *J. Chem. Phys.*, 114(12):5149–5155, 2001.

[Zim03] Urs Zimmerli, Michele Parrinello, and Petros Koumoutsakos. Dispersion corrections to density functional theory for water aromatic interaction. *Submitted to J. Chem. Phys.*, 2003.

[WWJ+03] T. Werder, J. H. Walther, R. L. Jaffe, T. Halicioglu, and P. Koumoutsakos. On the water-graphite interaction for use in MD simulations of graphite and carbon nanotubes. *J. Phys. Chem. B*, 107:1345–1352, 2003.

[Hal92] Thomas A. Halgren. Representation of van der Waals (vdW) interactions in molecular mechanics force fields: Potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.*, 114:7827–7843, 1992.

[Bon64] A. Bondi. van der Waals volumes and radii. *Parallel Computing*, 68(3):441–451, 1964.

[Bec93] Axel D. Becke. Density-functional thermochemistry. III. the role of exact exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.

[Lee88] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37(2):785–789, 1988.

[Per91] John P. Perdew. Unified theory of exchange and correlation beyond the local density approximation. In P. Ziesche and H. Eschrig, editors, *Electronic Structure of Solids*, pages 11–20. Akademie Verlag, Berlin, 1991.

[Per96] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865–3868, 1996.

[Bec88] A. D. Becke. Density-functional exchange-energy approximation with correct asymtotic behavior. *Phys. Rev. A*, 38(6):3098–3100, 1988.

[Boe00] A. Daniel Boese, Nikos L. Doltsinis, Ncholas C. Handy, and Michiel Sprik. New generalized gradient approximation functionals. *J. Chem. Phys.*, 112(4):1670–1678, 2000.

[CPMD01] CPMD V3.7. © IBM Corp. 1990-2003, © MPI für Festkörperforschung Stuttgart 1997-2001.

[Tro91] N. Troullier and José Luìs Martins. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B*, 43(3):1993–2006, 1991.

[Kle82] Leonard Kleinman and D. M. Bylander. Efficacious form for model pseudopotentials. *Phys. Rev. Lett.*, 48(20):1425–1428, 1982.

[Gau98] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople. Gaussian 98, revision a.7. Technical report, Gaussian, Inc., Pittsburgh PA, 1998.

[IAPWS01] The International Association for the Properties of Water and Steam. Guideline on the use of fundamental physical constants and basic constants of water. pages 1–7, 2001.

[Hin93] Alan Hinchliffe and Humberto J. Soscun Machado. Ab initio study of the static dipole polarizability of naphthalene. *Chem. Phys. Lett.*, 214(1):64–68, 1993.

[Ton91] Russell G. Tonkyn, Ralph Wiedmann, Edward R. Grant, and Michael G. White. Rotationally resolved photoionization of H_2O. *J. Chem. Phys.*, 95(10):7033–7040, 1991.

[Sch01] Detlef Schröder, Jessica Loos, Helmut Schwarz, Roland Thissen, Dorin V. Preda, Lawrence T. Scott, Doina Caraiman, Maxim V. Frach, and Diethard K. Böhme. Single and double ionization of corannulene and coronene. *Hel. Chim. Acta*, 84:1625–1634, 2001.

Flow of Aqueous Solutions in Carbon Nanotubes

S. C. Kassinos[1], J. H. Walther[2], E. Kotsalis[3], and P. Koumoutsakos[4]

[1] Dept. of Mechanical and Manufacturing Engineering, University of Cyprus, kassinos@ucy.ac.cy
[2] Institute of Computational Science, ETH Zurich, walther@inf.ethz.ch
[3] Institute of Computational Science, ETH Zurich, kotsalie@inf.ethz.ch
[4] Institute of Computational Science, ETH Zurich, petros@inf.ethz.ch

Summary. We conduct simulations of water flowing inside carbon nanotubes using non-equilibrium molecular dynamics simulations. A new adaptive forcing scheme is proposed to enforce a mean center of mass velocity. This scheme is compared to the non-adaptive, constant body/gravity force for the flow of methane in a carbon nanotube. The two schemes produce similar streaming velocity profiles and practically identical slip lengths. The wall slip predicted in the present simulations is found to be considerably shorter than the one observed in earlier studies using the constant body/gravity force scheme at considerably lower pressures. This observation is reminiscent of the slip length reduction with pressure increase that has been observed in the case of Couette flow of water. For water flowing through carbon nanotubes with diameters of 2.712, 4.068 and 5.424 nmand with flow speeds of $100\,\mathrm{m\,s^{-1}}$ we find slip lengths of 10, 12, and 18 nm. In addition, we consider mixtures of nitrogen and water flowing in a (20,20) carbon nanotube with diameter of 2.712 nm. For the mixture we find that the slip length is reduced to 6 nm as compared to 10 nm slip for the pure water. The shorter slip length is attributed to the fact that nitrogen forms droplets at the carbon surface, thus partially shielding the bulk flow from the hydrophobic effect.

1 Introduction

The transport of fluids through carbon nanotubes is important in many applications at the interface of bio- and nanotechnology and as a result it has been the subject of intense interest. For example, it has been proposed to use bundles of carbon nanotubes as artificial membranes where individual nanotubes will act as aquaporins, channels that in biological systems act to control water content in organs such as the red blood cells or the kidneys [KGH03, ZS03]. Carbon nanotube networks can also be used for the transport of water in nanofluidic devices such as lab-on-the-chip probes [PWW01], for molecular filtration membranes [KHM97] and in fuel cells. A range of CNT diameters and lengths are envisioned in these applications. CNTs can be manufactured in various sizes, ranging from several angstroms (Å) in diameter to more than 100 nm. It is therefore important to understand the physics of water flow through carbon nanochannels of different sizes. In these applications, the flow

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

of water is described remarkably well by macroscopic hydrodynamics provided the proper boundary conditions can be prescribed for the fluid-carbon interface. These hydrodynamic boundary conditions result from molecular interactions at the interface and are therefore affected by the degree of hydrophobicity of the CNT wall, possible doping of the inner CNT wall and the presence of impurities or solubles in the liquid phase. The ability to extract quantitative information for molecular dynamical (MD) simulations, for many of these applications, depends critically on being able to prescribe appropriate boundary conditions. For example, in MD simulations of water conduction through CNTs, water penetration and occupancy of the hydrophobic channel is sensitive to the solid-liquid interaction potentials. Relatively small changes in these potentials can lead to abrupt oscillations between dry and wet states [HRN01]. Such oscillations were linked to transitions between liquid and vapor phases of water [LCW99].

Navier-Stokes modeling of problems in nano-scale fluid mechanics may prove a computationally cost effective alternative to molecular dynamics simulations usually employed at these length scales [KB95] provided the complex fluid-solid interactions can find a suitable macroscopic model. At hydrophobic interfaces these interactions typically result in strong density fluctuations, anisotropic orientation of the water molecules [LMR84], and for solids with a high density, a finite fluid velocity (ΔU) at the interface [HvP60, Sch56, CSS84, BCTM01]. Thus, while the kinematic boundary condition of impermeability follows naturally from the definition of a fluid-solid interface, the issues relating to momentum transfer at the interface determining the dynamics of the problem is less clear. At moderate shear rates ($\partial u/\partial y$) the fluid remains Newtonian [LH89], and the slip velocity may be described by the linear relation:

$$\Delta U = L_s \frac{\partial u}{\partial y}, \tag{1}$$

where L_s is the slip length. Experimental evidence of slip has been demonstrated in studies of water in hydrophobized quartz capillaries [CSS84] and in drainage experiments [BCTM01] with slip lengths of $30 \pm 10\,\mathrm{nm}$ and $38 \pm 2\,\mathrm{nm}$, respectively. While most experiments have focused on the presence of slip at hydrophobic surfaces and on the possible validity of the no-slip condition at hydrophilic surfaces, recent colloid probe experiments of water on mica and glass have indicated a persistent slip of 8–9 nm at these hydrophilic surfaces [BKB02]. Molecular dynamics simulations of Poiseuille flow [BB99, TTE97, TG00] and planar Couette flow [TR90, TT97, CKB01] of simple Lennard-Jones fluids confined between Lennard-Jones solids have demonstrated the presence of both slip, no-slip and locking (negative slip length) depending on the "corrugation" of the surface. Thus, no-slip and locking are observed for low density solids, and slip is found to occur at strongly non-wetting interfaces, or for high density solids. In a recent study of methane confined between dense, graphite surfaces Sokhan *et al.* [SNQ01] found the slip to be fairly insensitive to the wetting properties of the surface, but to depend strongly on the density of the solid.

Many applications envision the use of large arrays of CNTs that will act either as sensors or as artificial, tunable, membranes. Molecular dynamics simulations of the

flow of simple fluids through a single CNT typically last for a few nanoseconds. In order to be able simulate the flow through entire arrays of nanotubes for long enough time intervals, so as to understand the operation of entire devices, one must resort to multiscale methods. For example, one could use a macroscopic flow solver to simulate the bulk of the fluid flow away from the CNTs, coupled through multiscale methods with a MD simulation of the detailed flow either inside individual CNTs or between closely spaced tubes. In the interest of simplifying the complexity of the overall problem, one might propose to use simplified equivalent continuum wall models that can be used in hybrid (molecular fluid, continuum wall) simulations of flow through nanochannels.

2 Physical Problem

In this study, we consider steady-state Poiseuille-like flow of methane, pure water, and a water-nitrogen mixture through a single-wall CNT. A constant flow rate is achieved through the use of a new adaptive forcing scheme that enforces a mean center-of-mass velocity. This scheme is found to compare favorably with the non-adaptive gravity-like force scheme [SNQ01]. Initially, fluid molecules are placed on regular lattice, and the tube is filled till a desired pressure level is achieved. Here pressure is measured as the radial component of the forces acting on carbon atoms of the nanotube per unit area. The flow is set to motion only after proper initial conditions have been achieved.

3 Governing equations and solution procedure

The present study employs nonequilibrium molecular dynamics (NEMD) simulations, where the carbon nanotube is treated as a rigid structure to allow the maximum time step of 2 fs imposed by the water model and to simplify the sampling. The influence of modeling the carbon nanotube as a rigid structure was recently found to be negligible for the static properties of water at carbon surfaces [WWJ+01, WWJ+03] and to be of secondary importance for dynamic properties such as the amount of slip [SNQ02].

3.1 Interaction potentials

The water is modeled using the rigid SPC/E water model by Berendsen *et al.* [BGS87], with bond lengths and angles constrained using SHAKE [vGB77]. The model involves a Lennard-Jones potential between the oxygen atoms

$$U_{\alpha\beta}(r_{ij}) = 4\epsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^{6}\right], \text{ for } r_{ij} < r_c, \tag{2}$$

where r_c is the radius of truncation ($U_{\alpha\beta}(r_{ij}) = 0$ for $r_{ij} > r_c$), and α and β refers to the atomic species (here oxygen-oxygen: $\epsilon_{\mathrm{OO}} = 0.6501\,\mathrm{kJ\,mol^{-1}}$ and $\sigma_{\mathrm{OO}} = 0.3166\,\mathrm{nm}$). The model furthermore involves a Coulomb potential acting between all atom pairs from different water molecules

$$U(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0}\frac{1}{r_{ij}}, \tag{3}$$

where ϵ_0 is the permittivity in vacuum, and q_i is the partial charge, $q_O = -0.8476$ and $q_H = 0.4238$, respectively [BGS87]. The present study uses a smooth truncation at $1.0\,\mathrm{nm}$ of the Coulomb interaction as described in Ref. [WJHK01].

The methane (CH_4) and nitrogen (N_2) molecules are treated as single Lennard-Jones sites with the parameters obtained from Refs. [SNQ02] and [BS87], thus $\epsilon_{\mathrm{N_2N_2}} = 0.2776\,\mathrm{kJ\,mol^{-1}}$, $\sigma_{\mathrm{N_2N_2}} = 0.3360\,\mathrm{nm}$, $\epsilon_{\mathrm{CH_4CH_4}} = 1.2310\,\mathrm{kJ\,mol^{-1}}$, $\sigma_{\mathrm{CH_4CH_4}} = 0.3810\,\mathrm{nm}$, respectively. Given these interaction potentials for the fluids, the wetting properties of the fluid-solid interface, and the ability of the interface to resist shear leading to a slip or no-slip condition, is determined by the potential function governing the fluid-solid interaction. For the water-carbon interaction we use a Lennard-Jones potential between the carbon and oxygen atoms of the water, with values from Werder *et al.* [WWJ+03] of $\epsilon_{\mathrm{CO}} = 0.4389\,\mathrm{kJ\,mol^{-1}}$, $\sigma_{\mathrm{CO}} = 0.3190\,\mathrm{nm}$. The corresponding parameters for CH_4 and N_2 are obtained by applying the Berthelot-Lorentz mixing rule, thus the N_2 parameters are $\epsilon_{\mathrm{N_2C}} = 0.2540\,\mathrm{kJ\,mol^{-1}}$, $\sigma_{\mathrm{N_2C}} = 0.3380\,\mathrm{nm}$, $\epsilon_{\mathrm{N_2O}} = 0.4248\,\mathrm{kJ\,mol^{-1}}$, $\sigma_{\mathrm{N_2O}} = 0.3263\,\mathrm{nm}$, and the CH_4 values are $\epsilon_{\mathrm{CH_4C}} = 0.5344\,\mathrm{kJ\,mol^{-1}}$, and $\sigma_{\mathrm{CH_4C}} = 0.3605\,\mathrm{nm}$, (see also Sokhan *et al.* [SNQ01]).

The governing Newton's equations are integrated in time using the leapfrog scheme subject to quasi-periodic boundary conditions in the direction along the axis of the carbon nanotube (z-axis) cf. Fig. 1.

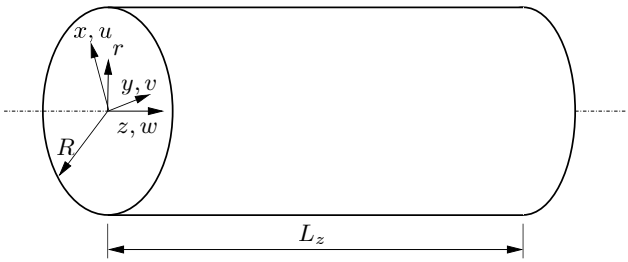


Fig. 1. Schematic of the system. R and z denote the radius of and the axis of the carbon nanotube. The size of the computational box along z is L_z.

3.2 Nonequilibrium Molecular Dynamics Simulations

The forcing of a mean flow in molecular dynamics simulations is usually achieved by imposing an acceleration ("gravity") on each fluid atom in the system. Typical values for the acceleration is 10^{11}–$10^{12}\,\mathrm{m\,s^{-2}}$ [SNQ01] depending on the amount friction exerted by the solid walls and the desired mean velocity. Thus, for the flow of methane in a carbon nanotube Sokhan *et al.* [SNQ02] used an acceleration of $0.5 \times 10^{12}\,\mathrm{m\,s^{-2}}$ to attain a mean velocity of $57\,\mathrm{m\,s^{-1}}$. However, this procedure may result in low frequency oscillations of the mean flow as observed by Sokhan *et al.* [SNQ01] for two-dimensional Poiseuille flow of methane between graphite surfaces. These oscillations with a wave length of $10\,\mathrm{ps}$, $0.2\,\mathrm{ns}$, and $2.0\,\mathrm{ns}$ extends the requirement for simulation time in order to sufficiently sample these slow modes. To improve the sampling rate, we propose an alternative adaptive scheme in which the imposed acceleration is computed dynamically to secure the desired (fixed) mean flow. Thus in a leapfrog approximation, the velocity of the molecules is updated according to

$$\boldsymbol{v}_i^{n+1/2} = \boldsymbol{v}_i^{n-1/2} + \frac{\delta t}{m_i}\left(\boldsymbol{f}_i + \boldsymbol{b}\right), \tag{4}$$

where δt is the time step, $\boldsymbol{v}_i$ and m_i the velocity and mass of the i-th molecule, and $\boldsymbol{f}_i$ and $\boldsymbol{b}$ denote the force and body force on the molecule, respectively. The center of mass velocity of the fluid molecules is updated accordingly

$$\boldsymbol{v}_{com}^{n+1/2} = \boldsymbol{v}_{com}^{n-1/2} + \frac{\delta t}{m_{tot}}\left(\boldsymbol{f}_{tot} + \boldsymbol{b}_{tot}\right), \tag{5}$$

where $\boldsymbol{f}_{tot}$ is the total force acting on the center of mass of the molecules, and $\boldsymbol{b}_{tot}$ is the total body force. In equation (4), $\boldsymbol{b}_{tot}$ is adjusted to yield the desired center of mass velocity ($\boldsymbol{v}_{com}^{n+1/2} = \boldsymbol{U}$), thus

$$\boldsymbol{b}_{tot} = \frac{m_{tot}}{\delta t}\left(\boldsymbol{U} - \boldsymbol{v}_{com}^{n-1/2}\right) - \boldsymbol{f}_{tot}. \tag{6}$$

In the following section we verify the ability of the present scheme to impose a mean flow by comparing the result from simulations of methane flowing in a carbon nanotube using a constant and the adaptive acceleration schemes.

4 Results

The simulations of fluids under confinement, such as the present study of dense liquids in carbon nanotubes, requires a careful selection of the number of molecules of the liquid to achieve the desired system pressure or the equivalent chemical potential. Thus, Sokhan *et al.* [SNQ02] performed grand canonical ensemble Monte Carlo simulations to determine the required number of fluid molecules. Here, we measure the pressure directly on the surface of the carbon nanotube during the equilibration of the system and adjust the number of molecules to obtain the desired pressure.

The ability of the fluid-solid interface to resist shear and the resulting macroscopic boundary condition is measured from the time average streaming velocity

$$\boldsymbol{u}_k = \frac{\sum_i^{n_k} m_i \boldsymbol{v}_i}{\sum_i^{n_k} m_i}, \tag{7}$$

where m_i and $\boldsymbol{v}_i$ are the mass and velocity of the i-th atom, and n_k denotes the number of atoms in the k-th bin. The statistics are sampled in polar bins with a radial bin resolution of $\approx 0.1\,\mathrm{nm}$ corresponding to 20–30 bins depending on the diameter of the carbon nanotube. The slip length (L_s) is extracted from the time average velocity profiles (7) using a least square fit to the parabola

$$W(r) = Ar^2 + Br + C. \tag{8}$$

Assuming $\partial W/\partial r|_{r=0} = 0$ gives $B = 0$ and thus a slip length of $L_s = \sqrt{-C/A} - R$. The fitting is performed on the data in the interval $r \in [0 : R - \sigma_{\alpha\beta}]$.

4.1 Flow of methane in carbon nanotubes

The flow of methane in carbon nanotubes was first studied by Sokhan *et al.* [SNQ02] who employed a constant acceleration to drive the flow. In the present study we repeat the simulation of methane in a (16,16) carbon nanotube using a constant acceleration and the new adaptive forcing. The diameter of the carbon nanotube is $2.1696\,\mathrm{nm}$ and the length of the computational box along the axis of the nanotube is $5.0\,\mathrm{nm}$ hosting a total of 132 methane molecules at a pressure of 67 bar. We first conduct a simulation with the adaptive scheme and an imposed mean velocity of $57\,\mathrm{m\,s^{-1}}$ corresponding to the mean flow obtained in Ref. [SNQ02] with a fixed acceleration of $0.5 \times 10^{12}\,\mathrm{m\,s^{-2}}$.

The methane molecules are initially placed on a regular lattice and the system is equilibrated for 20 ps, coupled to a Berendsen thermostat to achieve a temperature of 300 K. During the equilibration the mean velocity is set to zero. The adaptive scheme is switched on after an additional 20 ps of simulation (at total simulation time of 40 ps), and the methane molecules are again coupled to a Berendsen thermostat but only in the plane normal to the flow direction (the x-y plane). The simulation is continued for 16 ns sampling the results every 100 steps from 6 to 16 ns with a total of 50000 samples. The computed mean acceleration from this simulation is $1.6 \times 10^{13}\,\mathrm{m\,s^{-2}}$. A snapshot from the simulation is shown in Fig. 2.

In the second simulation we impose this constant mean acceleration throughout the simulation. The streaming velocity profile from the two simulations are compared in Fig. 3 demonstrating the ability of the adaptive scheme to force the flow. The extracted slip length for both cases is $\mathcal{O}(\sigma_{\mathrm{CH_4O}})$ cf. Fig. 4 and thus significantly smaller than the $6.4\,\mathrm{nm}$ measured by Sokhan *et al.* [SNQ02]. However, the elevated pressure used in the present study of 67 bar may result in a reduced slip as demonstrated by Walther *et al.* [WJW$^+$02] for the Couette flow problem with water between graphite surfaces.

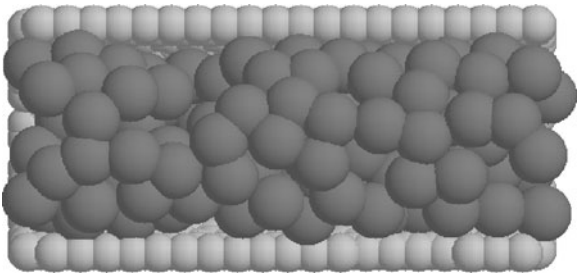


Fig. 2. Snapshot from NEMD simulations of the flow of methane inside a (16,16) carbon nanotube. Half of the carbon nanotube has been removed to view the methane molecules. (See Plate 18 on Page 277)

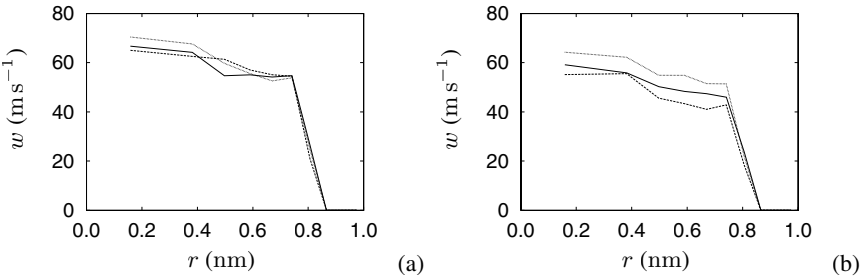


Fig. 3. The velocity profile from NEMD simulations of the flow of methane inside a (16,16) carbon nanotube. The adaptive scheme (a) result in less fluctuation on the time average velocity profile than the simulation using a constant acceleration (b).

4.2 The flow of water in carbon nanotubes

Next, we consider the flow of water through carbon nanotubes of different chiral vectors: (20,20), (30,30) and (40,40), corresponding to tube diameters of 2.712, 4.068, and 5.424,nm, respectively. The mean flow speed is $100\,\mathrm{m\,s^{-1}}$ and the pressure in the system is in the range 400–600 bar sufficiently high to secure a complete filling of the tube with a total number of water molecules of 736, 904, and 1704, respectively. The simulations are conducted for 2 ns sampling the streaming velocity profiles from 1–2 ns with a total of 5000 snapshots. The slip lengths extracted from the streaming velocity profiles shown in Fig. 5 are 10, 12, and 18 nm, for the 2.712, 4.068, and 5.424 nm diameter tubes, respectively. Thus, the flow in the carbon nanotube with the smallest radius displays the smallest slip length, consistent with increasing surface friction as the radius is reduced. The observed values are furthermore in agreement with the experimental value of 30 nm for water at an extended hydrophobic surface [CSS84].

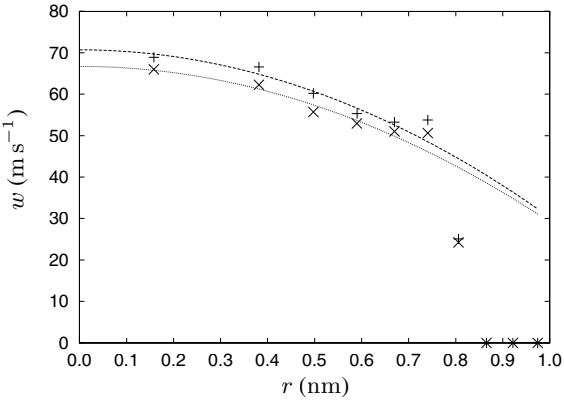


Fig. 4. The velocity profile from NEMD simulations of the flow of methane inside a (16,16) carbon nanotube. Both the adaptive scheme (+, - -) and the constant acceleration (×, —) predict a slip length of $\mathcal{O}(\sigma_{\mathrm{CH_4O}})$.

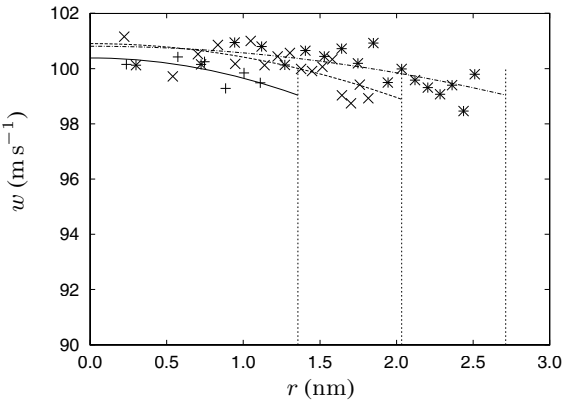


Fig. 5. Time average streaming velocity profiles of water in a carbon nanotubes of different diameter: 2.712 nm: (+, —); 4.068 nm: (×, – –); 5.424 nm: (∗, –·–). The points indicate the values extracted from the NEMD simulations and the lines a least squared fit to Eq. (8). The vertical lines mark the position of the surface of the carbon nanotubes.

4.3 Flow of water-nitrogen mixtures in carbon nanotubes

The effect of gas bubbles in hydrophobic systems is believed to influence the static properties of the system e.g., hydration [AMT02], but also dynamic properties such as the amount of slip experienced at the fluid-solid interface [Vin99]. To study these effects, we consider a mixture of water and air flowing inside a (20,20) carbon nanotube. The air here is represented by nitrogen molecules dissolved in the water. At 300 K and 600 bar the employed molar concentration of N_2 of 0.075 M is supersaturated and corresponds to 50 nitrogen and 670 water molecules.

The forcing of the mean flow and the thermostat is only applied to the water molecules, which drives the system hydrodynamically and thermally. The nitrogen is initially placed randomly within the water as shown in Fig. 6a. During the equilibration of the system the nitrogen molecules desorp from the water phase forming high density, supercritical droplets at the carbon surface. The time average streaming velocity extracted from the trajectory in the time interval 1–2 ns is compared with the result from the pure water simulation cf. Fig. 7. The extracted slip length is 6 nm approximately 4 nm less than slip experienced for pure water. The observed decrease in slip length is attributed to the fact that nitrogen forms supercritical droplets at the carbon surface, thus partially shielding the bulk flow from the hydrophobic effect.

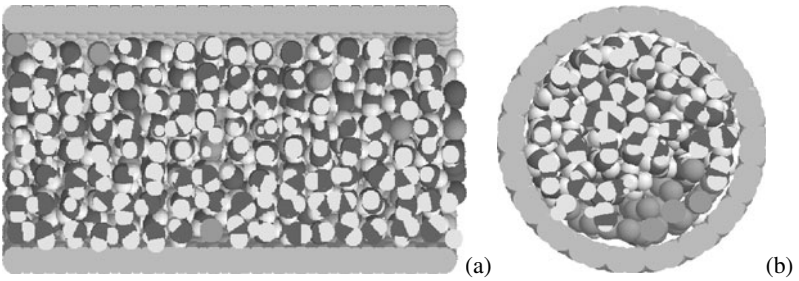


Fig. 6. Snapshots of the initial molecular configuration (a) and the equilibrated system (b) of a water-nitrogen mixture flowing inside a carbon nanotube. (See Plate 19 on Page 277)

5 Summary

We have presented nonequilibrium molecular dynamics simulations of the flow of methane, water, and a water-nitrogen mixture inside carbon nanotubes. In these simulations we implemented a new adaptive scheme for imposing a mean fluid velocity through the system. We compared the adaptive scheme to the non-adaptive imposition of a constant body/gravity force for the flow of methane in a carbon nanotube. We found that the adaptive scheme is as effective in imposing the selected mean

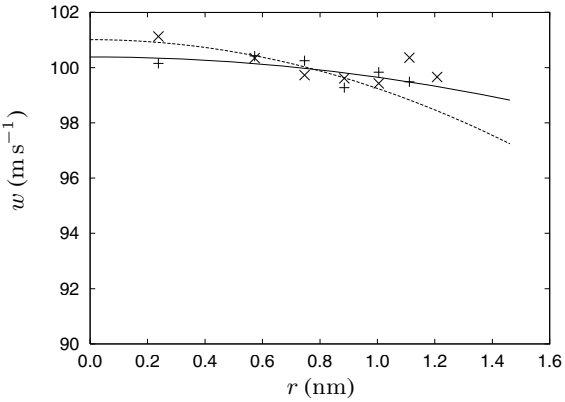


Fig. 7. Time average streaming velocity profiles of a water-nitrogen mixture ($\times$, - -) in a carbon nanotube. The slip length is approximately $6\,\mathrm{nm}$, and $4{,}\mathrm{nm}$ smaller than the corresponding results for pure water ($*$, —).

velocity, but seems to be free of the oscillations that are often produced by the non-adaptive scheme. The slip lengths observed in these simulations of methane through carbon nanotubes are much smaller than what has been observed in earlier studies. The difference is ascribed to the higher pressure in our simulations as compared to the pressure reported in the earlier studies.

We also considered water and a mixture of water and nitrogen flowing through carbon nanotubes with various diameters. For the case of pure water, we found the slip length to increase monotonically with tube diameter. For the case of the mixture, we found the slip length to be slightly shorter than what was observed in the flow of pure water through the same carbon nanotube. We suspect that the shorter slip length in the case of the mixture is due to the formation of supercritical nitrogen droplets at the carbon surface that partially shields the bulk flow from the hydrophobic carbon surface.

Acknowledgments

We wish to thank Richard L. Jaffe for critically reading the manuscript.

References

[AMT02] P. Attard, M. P. Moody, and J. W. G. Tyrrell. Nanobubbles: the big picture. *Physica A*, 314(1-4):696–705, 2002.

[BB99] Jean-Louis Barrat and Lydéric Bocquet. Large slip effect at a nonwetting fluid-solid interface. *Phys. Rev. Lett.*, 82(23):4671–4674, 1999.

[BCTM01] J. Baudry, E. Charlaix, A. Tonck, and D. Mazuyer. Experimental evidence for a large slip effect at a nonwetting fluid-solid interface. *Langmuir*, 17:5232–5236, 2001.

[BGS87] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91:6269–6271, 1987.

[BKB02] Elmar Bonaccurso, Michael Kappl, and Hans-Jürgen Butt. Hydrodynamic force measurements: Boundary slip of water on hydrophilic surfaces and electrokinetic effects. *Phys. Rev. Lett.*, 88(7):076103, 2002.

[BS87] Mary J. Bojan and William A. Steele. Interactions of diatomic molecules with graphite. *Langmuir*, 3(6):1123–1127, 1987.

[CKB01] Marek Cieplak, Joel Koplik, and Jayanth R. Banavar. Boundary conditions at a fluid-solid interface. *Phys. Rev. Lett.*, 86(5):803–806, 2001.

[CSS84] N. V. Churaev, V. D. Sobolev, and A. N. Somov. Slippage of liquids over lyophobic solid surfaces. *J. Coll. Interface Sci.*, 97(2):574–581, 1984.

[HRN01] G. Hummer, J. C. Rasaiah, and J. P. Noworyta. Water conduction through the hydrophobic channel of a carbon nanotube. *Nature*, 414:188–190, 2001.

[HvP60] H. Helmholtz and G. von Piotrowski. Über reibung tropfbarer flussigkeiten. *Sitzungsberichte der Kaiserlich Akademie der Wissenschaften*, 40:607–658, 1860.

[KB95] Joel Koplik and Jayanth R. Banavar. Continuum deductions from molecular hydrodynamics. *Annu. Rev. Fluid Mech.*, 27:257–292, 1995.

[KGH03] Amrit Kalra, Shekhar Garde, and Gerhard Hummer. Osmotic water transport through carbon nanotube membranes. *Proc. Natl. Acad. Sci. USA*, 100(18):10175–10180, 2003.

[KHM97] B. Jirage Kshama, John C. Hulteen, and Charles R. Martin. Nanotube-based molecular-filtration membranes. *Science*, 278(5338):655–658, 1997.

[LCW99] Ka Lum, David Chandler, and John D. Weeks. Hydrophobicity at small and large length scales. *J. Phys. Chem. B*, 103:4570–4577, 1999.

[LH89] W. Loose and S. Hess. Rheology of dense model fluids via nonequilibrium molecular dynamics: shear thinning and ordering transition. *Rheol. Acta*, 28:91–101, 1989.

[LMR84] C. Y. Lee, J. A. McCammon, and P. J. Rossky. The structure of liquid water at an extended hydrophobic surface. *J. Chem. Phys.*, 80(9):4448–4455, 1984.

[PWW01] M. W. J. Prins, W. J. J. Welters, and J. W. Weekamp. Fluid control in multichannel structures by electrocapillary pressure. *Science*, 291:277–280, 2001.

[Sch56] Erhard Schnell. Slippage of water over nonwettable surfaces. *J. Appl. Phys.*, 27(10):1149–1152, 1956.

[SNQ01] V. P. Sokhan, D. Nicholson, and N. Quirke. Fluid flow in nanopores: An examination of hydrodynamic boundary conditions. *J. Chem. Phys.*, 115(8):3878–3887, 2001.

[SNQ02] Vladimir P. Sokhan, David Nicholson, and Nicholas Quirke. Fluid flow in nanopores: Accurate boundary conditions for carbon nanotubes. *J. Chem. Phys.*, 117(18):8531–8539, 2002.

[TG00] Karl P. Travis and Keith E. Gubbins. Poiseuille flow of Lennard-Jones fluids in narrow slit pores. *J. Chem. Phys.*, 112(4):1984–1994, 2000.

[TR90] Peter A. Thompson and Mark O. Robbins. Shear flow near solids: Epitaxial order and flow boundary conditions. *Phys. Rev. A*, 41(12):6830–6841, 1990.

[TT97] Peter A. Thompson and Sandra M. Troian. A general boundary condition for liquid flow at solid surfaces. *Nature*, 389:360–362, 1997.

[TTE97] Karl P Travis, B. D. Todd, and Denis J. Evans. Departure from Navier-Stokes hydrodynamics in confined liquids. *Phys. Rev. E*, 55(4):4288–4295, 1997.

[vGB77] W. F. van Gunsteren and H. J. C. Berendsen. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.*, 37(5):1311–1327, 1977.

[Vin99] Olga I. Vinogradova. Slippage of water over hydrophobic surfaces. *Int. J. Miner. Process.*, 56:31–60, 1999.

[WJHK01] J. H. Walther, R. Jaffe, T. Halicioglu, and P. Koumoutsakos. Carbon nanotubes in water: Structural characteristics and energetics. *J. Phys. Chem. B*, 105:9980–9987, 2001.

[WJW+02] J. H. Walther, R. Jaffe, T. Werder, T. Halicioglu, and P. Koumoutsakos. On the boundary condition for water at a hydrophobic surface. In *Proceedings of the Summer Program 2002*, pages 317–329, Center for Turbulence Research, Stanford University/NASA Ames, 2002.

[WWJ+01] T. Werder, J. H. Walther, R. Jaffe, T. Halicioglu, F. Noca, and P. Koumoutsakos. Molecular dynamics simulations of contact angles of water droplets in carbon nanotubes. *Nano Letters*, 1(12):697–702, 2001.

[WWJ+03] T. Werder, J. H. Walther, R. L. Jaffe, T. Halicioglu, and P. Koumoutsakos. On the water-graphite interaction for use in MD simulations of graphite and carbon nanotubes. *J. Phys. Chem. B*, 107:1345–1352, 2003.

[ZS03] Fangquian Zhu and Klaus Schulten. Water and proton conduction through carbon nanotubes as models for biological channels. *Biophys. J.*, 85(1):236–244, 2003.

Continuum-particle hybrid methods for dense fluids

Thomas Werder, Jens H. Walther, Joonas Asikainen, and Petros Koumoutsakos

Institute of Computational Science, ETH Zürich, CH-8092 Zürich, Switzerland.
{werder,walther,tommia,koumoutsakos}@inf.ethz.ch

Summary. We outline a hybrid multiscale algorithm for the simulation of dense fluids by coupling molecular dynamics and continuum fluid dynamics, here described by the incompressible Navier-Stokes equations. We estimate the required sampling of the atomistic system to achieve sufficiently accurate boundary condition for the continuum system. This analysis indicate that momentum flux (pressure) tensor requires an intractable number of samples ($\mathcal{O}(10^8)$), hence penalizing flux based coupling algorithms. For the present systems we therefore employ a density based, alternating Schwarz algorithm to couple the two system. The boundary conditions at the interface between the two regions is imposed by inserting or deleting atoms in a buffer region using the USHER algorithm. Preliminary results indicate, that the forcing of the continuum may be achieved using appropriate source terms in the momentum equations.

1 Introduction

The study of nanoscale fluid mechanics [KZWW03] is of key importance for the understanding and development of biosensors. From the computational side, classical molecular dynamics have so far been the main tool to characterize nanoscale fluid mechanics phenomena such as wetting and hydrophobicity. A canonical system of particular interest with regard to biosensors is the interaction of carbon nanotubes with water. This system has recently been studied in some detail, see Refs. [GM03, WWJ$^+$03, WWJK03] and references therein. Simulations of CNTs in aqueous environments using molecular dynamics are limited to volumes of the order $O(10\,\mathrm{nm}^3)$ and time scales of $O(10\,\mathrm{ns})$. In order to develop computational tools that will assist in the design of biosensors a multiscale approach is indispensable.

For the case of dilute particle systems, there exist sophisticated coupling methods as described and reviewed by Garcia *et al.* [GBCA99]. For dense fluids, two classes of coupling schemes have been proposed which all split the computational domain into a particle region (P) and a continuum (C) region that are connected through an overlap region (X). The two classes of coupling schemes differ in that the first one is based on direct flux exchange [OT95, FWF00, DBC03a] and the second one on

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

the Schwarz alternating method [HP97, Had99]. The usage of the Schwarz method avoids the imposition of fluxes in X since flux continuity is automatically ensured if the transport coefficients in the two regions are consistent.

Here, we first describe an algorithm that imposes the boundary conditions from the continuum to the particle system. This algorithm is tested on a two-dimensional Lennard-Jones fluid, and its extension to three-dimensional systems is discussed. Second, we describe the coupling of the molecular dynamics code FASTTUBE [WWJK03] to the commercial flow solver STARCD [sta99] for the case of three dimensional incompressible flows.

2 Imposing boundary conditions on the MD

2.1 Flux based coupling vs. density based coupling

Flux based schemes [FWF00, DBC03a] are appealing because they conserve mass, momentum, and energy by construction and they can cope with small differences in the transport coefficients of the two descriptions [DBC03a]. Despite this, we propose a hybrid scheme that is based on the mutual imposition of the mass density ρ, of the mean velocity $\boldsymbol{u}$, and of the temperature T. Such a scheme has the advantage that it decouples time scales, that densities are more easily sampled than fluxes, and that it is straightforward to implement in commercial codes.

Recently, Hadjiconstantinou *et al.* [HGBH03] derived an estimate for the number of independent samples needed in molecular systems to obtain a fractional error E_q in a quantity q that is measured in a domain of volume V. The fractional error of a quantity q is defined as $E_q = \sigma(\bar{q})/\bar{q} = \sigma(q)/\sqrt{M}\bar{q}$, where $\sigma(q)$ and $\bar{q}$ denote the standard deviation and average of q, and M is the number of statistically independent samples. The expressions for the mean velocity, the density, and the pressure are [HGBH03]

$$M_u = \frac{k_B T_0}{\bar{u}_x^2} \frac{1}{\rho V {E_u}^2}, \quad M_\rho = \frac{\kappa_T k_B T_0}{V {E_\rho}^2}, \quad M_P = \frac{\gamma k_B T_0}{P_0^2 \kappa_T} \frac{1}{V {E_P}^2}. \tag{1}$$

These expressions evaluated for a $1\,\mathrm{nm}^3$ bulk water cell at atmospheric conditions (pressure $P = 1\mathrm{bar}$, temperature $T = 293\,\mathrm{K}$, ratio of specific heats $\gamma = 1$, isothermal compressibility of $\kappa_T = 48.95 \cdot 10^{-6}\,\mathrm{bar}^{-1}$), and with a mean velocity $\boldsymbol{u} = 100\,\mathrm{ms}^{-1}$ yield a number of required samples for a fractional error of 5% of $M_u \approx 10^2$, $M_\rho \approx 1$, and $M_P \approx 10^8$ for velocity, density, and pressure, respectively. We emphasize that these are the numbers of *statistically independent* samples, i.e., the number of time steps needed to achieve this sampling is substantially higher, and that the volume of a $1\,\mathrm{nm}^3$ represents a realistic cell size. The prohibitive sampling costs to measure the momentum flux (pressure) tensor favor the use of a density based scheme. A possible alternative solution to this problem is to use a mesoscopic regime between the microscopic and macroscopic domains, where the fluid flow is described using fluctuating hydrodynamics (FH) [LL59].

2.2 Overview of the hybrid algorithm

The algorithm is based on the Schwarz iteration scheme [Had99], i.e., the computational domain consists of pure continuum cells C, pure particle cells P, and overlap cells X, where both descriptions are valid. One full iteration cycle is described as:

Step 1) Provide an initial guess for the solution in the C domain.
Step 2) Run an MD simulation until a steady state is reached, with the continuum state $(\rho, \boldsymbol{u}, T)$ imposed to the boundary cells ($C \rightarrow P$) of the atomistic system.
Step 3) Solve the continuum equations with the steady state solution of the MD as boundary conditions in the $P \rightarrow C$ cells.
Step 4) If not converged, return to Step (2).

The two main difficulties in coupling molecular and continuum systems are the introduction and removal of particles and the altered dynamics of the molecules in the $C \rightarrow P$ region due to the "missing particles" exteriour to the molecular domain. The first problem is solved using the USHER algorithm [DBC03b] and for the second problem, we extend the notion of stochastic boundary condition introduced to simulate the contact of an MD system with a thermal reservoir [CT80] to cover systems with a non-vanishing mean velocity across the boundary.

A generic molecular dynamics time step, where $(\rho, \boldsymbol{u}, T, \boldsymbol{\nabla u}, \boldsymbol{\nabla} T)$ are given in the $C \rightarrow P$ cells, reads:

1. *Force computation.* Compute forces according to the pair potential U.
2. *Velocity update.* Update the particle velocities using the leap-frog scheme.
3. *Andersen thermostat.* Choose particles randomly in the $C \rightarrow P$ cells with probability ν. Reset their velocities according to a biased Maxwell-Boltzmann distribution ($f(\widetilde{u}_k)$) with mean velocity $\boldsymbol{u}'(\Delta \boldsymbol{x}) = \boldsymbol{u} + \boldsymbol{\nabla u}^T \Delta \boldsymbol{x}$ and temperature $T'(\Delta \boldsymbol{x}) = T + \boldsymbol{\nabla} T^T \Delta \boldsymbol{x}$:

$$f(\widetilde{u}_k) = \sqrt{\frac{m}{2\pi k_b T'}} \exp\left(-\frac{(\widetilde{u}_k - u')^2}{2k_b T'}\right). \tag{2}$$

   where (u', T') denote the new thermostatted state, and $\widetilde{u}_k$ the instantaneous velocity of the k-th particles.
4. *Position update.* Update the particle positions using the leap-frog scheme. The particles in the $C \rightarrow P$ cells are updated in three steps, first with the mean velocity subtracted, second, possible collisions with stochastic walls are evaluated (see next step), and third the displacement due to the mean velocity is added.
5. *Stochastic wall.* The boundary between any $C \rightarrow P$ and the C cells, is formed by a stochastic wall W [CT80] located at $\boldsymbol{x}_{W'} = \boldsymbol{x}_W - \delta t(\boldsymbol{u} \cdot \boldsymbol{n}_\perp)$ and with outward normal $\boldsymbol{n}_\perp$. The normal velocity of particles k that hit W is sampled from the probability density [TTKB98]

$$f(\widetilde{u}_k) = \frac{m}{k_b T'} |u_k| \exp\left(-\frac{(\widetilde{u}_k - u')^2}{2k_b T'}\right), \tag{3}$$

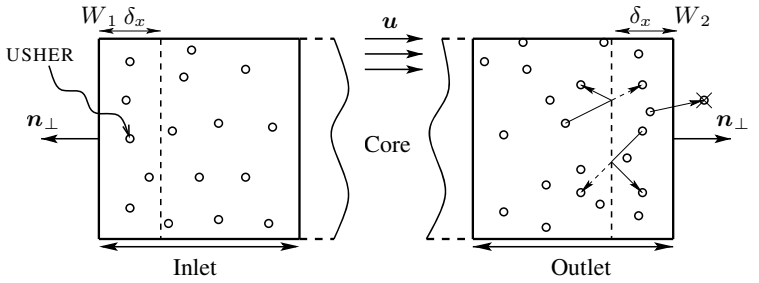


Fig. 1. Imposition of $(\rho, \boldsymbol{u}, T)$ in the $C \to P$ cells. An Andersen thermostat imposes $\boldsymbol{u}$ and T. For $\boldsymbol{u} \neq \boldsymbol{0}$, the stochastic walls W_i' are shifted by $\delta_x = |\boldsymbol{u} \cdot \boldsymbol{n}_\perp|\delta t$, where $\boldsymbol{n}_\perp$ is the outward normal, and the collisions with the wall are evaluated with the mean velocity subtracted. The USHER algorithm [DBC03b] is used to insert particles in the region δx at a rate of $\rho h_y \boldsymbol{u} \cdot \boldsymbol{n}_\perp$, where h_y denote the height of the cell.

while the tangential components are sampled from Eq. (2). Finally, add the mean displacement $\boldsymbol{x}_k^{t+1} = \boldsymbol{x}_k^t + \delta t_p \boldsymbol{u}$.

6. *Mass flux.* The mass flux per time step across W is given by $\delta_m = \delta t \rho h_y \boldsymbol{u} \cdot \boldsymbol{n}_\perp$, i.e., at the inlet, every m/δ_m steps, a particle is inserted using the USHER algorithm [DBC03b] in a region of width δ_x, cf. Fig. 1.
7. If no steady state is reached, return to step 1 with $t \leftarrow t + 1$.

2.3 Numerical test

We consider a two dimensional 6–12 Lennard-Jones fluid with interaction potential $U(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]$ and use non-dimensionalized units of length (σ), energy (ϵ), and time $(m\sigma^2/48\epsilon)^{1/2}$. We employ a cutoff radius of 3 and the leap-frog integration algorithm with time step 0.002. First, we study a system in thermodynamic equilibrium with zero mean velocity, with density 0.64 (1474 particles), and with a temperature of 1.0. Second, a parallel flow is imposed along the x-axis, i.e. the mean velocity is $\boldsymbol{u} = (1.0, 0.0)$. And third, a Couette flow is imposed with a velocity at the left of $\boldsymbol{u}_l = (0.0, -1.0)$, and $\boldsymbol{u}_r = (0.0, 1.0)$ at the right boundary.

The computational box is of size 48×48 and the $C \to P$ cells have an edge length of 6. The density, temperature and velocity are computed in bins of size 1×48 and time averaged using every 100th sample between the steps 10^5 and $5 \cdot 10^5$ of the production run. For the Andersen thermostat, we use a collision frequency of $\nu = 0.01$. In Fig. 2, density and temperature profiles are given for all four cases. For the equilibrium case, the density increases near the domain boundaries due to the "missing" particles on the outside of the domain. This effect remains however localized and does not affect the performance of the overall algorithm [CT80]. The Couette flow displays the expected parabolic temperature profile [ZGA02], and linear velocity profile, cf. Figs. 2b and 2d.

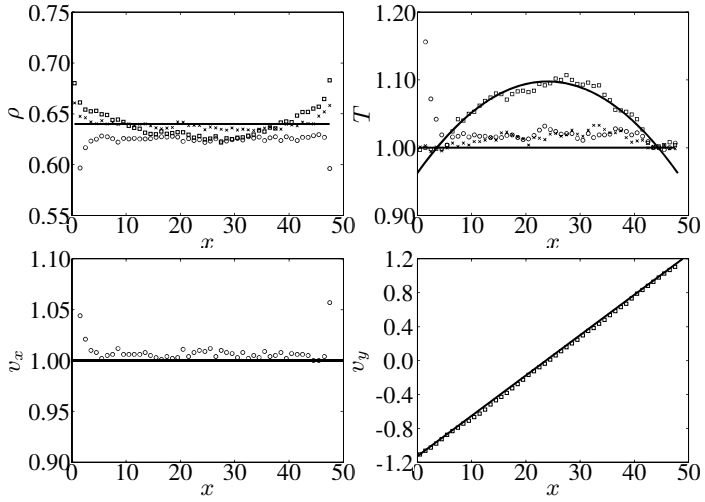


Fig. 2. Upper row: Density and temperature profiles along the x-axis for the equilibrium (crosses), the parallel flow (circles), and the Couette flow (squares). The straight lines are at constant density (0.64), constant temperature (1.0), and a parabolic fit to the temperature profile for the Couette flow. Lower row: Velocity profiles for the parallel flow (v_x), and the Couette flow (v_y). The straight lines are at $v_x = 1.0$ (left) and a linear fit to v_y (right).

3 Imposing boundary conditions on the continuum system

To study the properties of the alternating Schwarz algorithm for the coupling of fully three-dimensional molecular and continuum systems we have interfaced a commercially available Navier-Stokes solver (StarCD [sta99]), with an in-house MD code (FASTTUBE [WJHK01]). We consider a molecular system that consists of bulk water, and for testing purposes water flowing past a carbon nanotube. The water is described by the SPC/E water model, and the carbon nanotube is modeled as a rigid structure interacting with the water through a Lennard-Jones potential. The continuum system is described by the isothermal, incompressible, Navier-Stokes (NS) equations.

3.1 Governing equations

The interaction potentials governing the molecular system and their numerical treatment is described in detail in Refs. [WJHK01, WWJK03, WWJ+03]. To simplify the treatment of the molecular system, periodic boundary conditions are imposed in all spatial directions. Thus, the dynamics of the unforced system follows that of a microcanonical system, leaving the particles in the $P \rightarrow C$ region in an equilibrated "state" which only requires a weak coupling to a Berendsen thermostat [BPvG+84]. Furthermore, the periodicity in the particle positions secures a global conservation of mass and avoids the need to add and delete particles. However, it effectively requires that the continuum fields are periodic at the interface to the molecular system,

limiting its general applicability. However, this allows us to study the coupling of the two systems and the imposition of boundary conditions on the continuum system.

The averaged molecular velocity in the $C \rightarrow P$ region is forced to match the continuum values by adding a body force ($\boldsymbol{b}_{tot}$) to the center of mass of the molecules in the $C \rightarrow P$ region

$$\boldsymbol{v}_{com}^{n+1/2} = \boldsymbol{v}_{com}^{n-1/2} + \frac{\delta t}{m_{tot}} \left(\boldsymbol{f}_{tot} + \boldsymbol{b}_{tot} \right), \tag{4}$$

where $\boldsymbol{v}_{com}$, m_{tot}, and $\boldsymbol{f}_{tot}$ denote the center of mass velocity, the total mass and force on the $C \rightarrow P$ region, and $\boldsymbol{b}_{tot}$ is the required body force such that $\boldsymbol{v}_{com}^{n+1/2} = \boldsymbol{U}$, where $\boldsymbol{U}$ is the continuum velocity field.

For the bulk water systems considered in the present study, the continuum approximation is assumed to be governed by the isothermal, incompressible Navier-Stokes equations. To couple the two systems, we overlay the molecular system with a finite-volume mesh covering the entire computational domain, as shown in Fig. 3. The size of the finite-volume cells is chosen to match the size and location of the bins used in the molecular system for the averaging and for the imposition of the boundary conditions. In the present study we use cells of size $0.5 \times 0.5 \times 1\,\mathrm{nm}$ to secure a sufficient averaging and forcing of the molecular system. By the complete overlay, we avoid explicit interfaces in the continuum system.

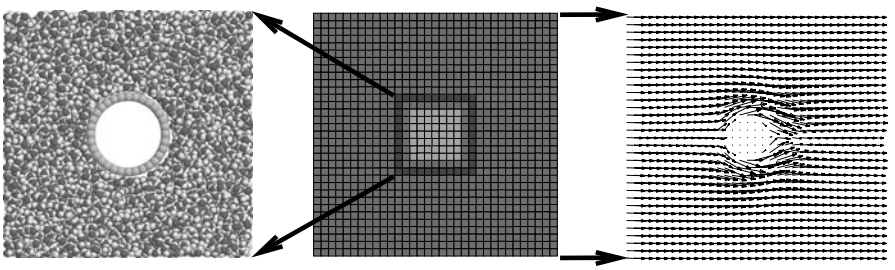


Fig. 3. The continuum mesh overlays the molecular system and covers the entire computational domain. The light gray area in (middle figure) indicates the $P \rightarrow C$ region, and the dark area the $C \rightarrow P$ region.

3.2 Results

To demonstrate the ability of the $P \rightarrow C$ forcing we first consider the flow of water past a carbon nanotube (Fig. 3). Walther *et al.* [WWJK03] have studied this system using non-equilibrium molecular dynamics and observed that the water exhibits a no-slip boundary condition. In the current hybrid formulation, the velocity field at the $C \rightarrow P$ interface is non-periodic, but nevertheless the simulation illustrates that

body forces can impose the required molecular field on the continuum, cf. Fig. 3. The exteriour boundary condition is periodic with a streaming velocity of $100\mathrm{ms}^{-1}$ in the x direction and zero mass flux in the y and z directions. The molecular system consists of approximately 13177 atoms, and requires 20000 time steps for each outer iteration to secure a proper average, and corresponds to approximately 9 h on an 1.4 GHz AMD Athlon processor. The continuum mesh consists of $33 \times 33 \times 1$ mesh points and a steady state is typically found in 100 time steps requiring less than 1 cpu minute. At a later stage in the simulation, the molecular system develops a vapor phase at the corners of the P system due to the improper boundary conditions. Utilizing the USHER algorithm for this system would resolve this problem.

The second test case involves a Couette flow with a molecular system of $(4 \times 4\,\mathrm{nm})$ located at the center of the computational domain $(16 \times 16\,\mathrm{nm})$, similar to the previous configuration. For this system the $C \rightarrow P$ BC are periodic, and the simulation reaches a steady state with a linear velocity profile extending through the molecular system, cf. Fig. 4. Notice, that both the molecular and continuum systems are forced with a Dirichlet BC for the velocity which suffices for the present linear velocity profile. However, a non-linear velocity profile (e.g., the Poiseuille flow) requires a specification of both the velocity and the gradient of the velocity – a boundary condition of the Robin type. This may be achieve by applying the Andersen thermostat as in the 2D study in Section 2 or by using a finite-difference approximation to the gradient by forcing two neighbouring cells.

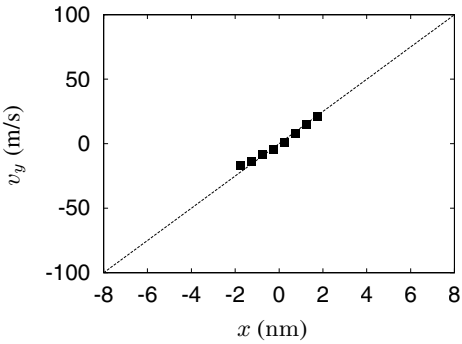


Fig. 4. The velocity profile from the Couette flow problem. + + +: time averaged molecular profile; —: continuum profile.

4 Summary

We have discussed parts of a hybrid algorithm for dense fluids described by molecular systems in one part of the computational domain and as continuum systems in another part. In particular, we have presented and tested an algorithm to perform molec-

ular dynamics simulations with non-periodic boundary conditions. Furthermore, we have conducted preliminary (periodic) simulations of the flow of water around a carbon nanotube, where the vicinity of the tube was modeled with molecular dynamics and the rest of the computational domain with a Navier-Stokes solver. Currently, we consider the use of fluctuating hydrodynamics to describe the continuum part of the algorithm, since this would allow for massive reductions in the computational cost for sampling. We plan to combine these elements in the near future in order to obtain a general algorithm for the multiscale simulation of dense fluids.

References

[BPvG+84] 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(8):3684–3684, 1984.

[CT80] G. Ciccotti and A. Tenenbaum. Canonical ensemble and nonequilibrium states by molecular dynamics. *J. Stat. Phys.*, 23(6):767–772, 1980.

[DBC03a] R. Delgado-Buscalioni and P. V. Coveney. Continuum-particle hybrid coupling for mass, momentum, and energy transfers in unsteady flow. *Phys. Rev. E*, 67:046704–1–046704–13, 2003.

[DBC03b] R. Delgado-Buscalioni and P. V. Coveney. USHER: An algorithm for particle insertion in dense fluids. *J. Chem. Phys.*, 119(2):978–987, 2003.

[FWF00] E. G. Flekkøy, G. Wagner, and J. Feder. Hybrid model for combined particle and continuum dynamics. *Europhys. Lett.*, 52(3):271–276, 2000.

[GBCA99] Alejandro L. Garcia, John B. Bell, William Y. Crutchfield, and Berni J. Alder. Adaptive mesh and algorithm refinement using direct simulation Monte Carlo. *J. Comput. Phys.*, 154:134–155, 1999.

[GM03] M. C. Gordillo and J. Martí. Water on the outside of carbon nanotube bundles. *Phys. Rev. B*, 67:205425–1–205425–4, 2003.

[Had99] N. G. Hadjiconstantinou. Hybrid atomistic-continuum formulations and the moving contact-line problem. *J. Comput. Phys.*, 154:245–265, 1999.

[HGBH03] Nicolas G. Hadjiconstantinou, Alejandro L. Garcia, Martin Z. Bazant, and Gang He. Statistical error in particle simluations of hydrodynamic phenomena. *J. Comput. Phys.*, 187:274–297, 2003.

[HP97] Nicolas G. Hadjiconstantinou and Anthony T. Patera. Heterogeneous atomistic-continuum representations for dense fluid systems. *Int. J. Mod. Phy. C*, 8(4):967–976, 1997.

[KZWW03] P. Koumoutsakos, U. Zimmerli, T. Werder, and J. H. Walther. Nanoscale fluid mechanics. *ASME/SPIE handbook of Nanotechnology*, page in press, 2003.

[LL59] L. D. Landau and E. M. Lifshitz. *Fluid Mechanics*, volume 6. Pergamon Press, Oxford, 1959.

[OT95] Sean T. O'Connell and Peter A. Thompson. Molecular dynamics-continuum hybrid computations: A tool for studying complex fluid flow. *Phys. Rev. E*, 52(6):R5792–R5795, 1995.

[sta99] STAR-CD version 3.1a. manual, 1999.

[TTKB98] Riina Tehver, Flavio Toigo, Joel Koplik, and Jayanth R. Banavar. Thermal walls in computer simulations. *Phys. Rev. E*, 57(1):R17–R20, 1998.

[WJHK01] J. H. Walther, R. Jaffe, T. Halicioglu, and P. Koumoutsakos. Carbon nanotubes in water: Structural characteristics and energetics. *J. Phys. Chem. B*, 105:9980–9987, 2001.

[WWJ+03] T. Werder, J. H. Walther, R. L. Jaffe, T. Halicioglu, and P. Koumoutsakos. On the water-graphite interaction for use in MD simulations of graphite and carbon nanotubes. *J. Phys. Chem. B*, 107:1345–1352, 2003.

[WWJK03] J. H. Walther, T. Werder, R. L. Jaffe, and P. Koumoutsakos. Hydrodynamic properties of carbon nanotubes. *Phys. Rev. E*, page accepted, 2003.

[ZGA02] Yihao Zheng, Alejandro L. Garcia, and Berni J. Alder. Comparison of kinetic theory and hydrodynamics for Poiseuille flow. *J. Stat. Phys.*, 109(3/4):495–505, 2002.

Dissipative Particle Dynamics for Modeling Complex Fluidics

Justyna Czerwinska[1] and Nikolaus A. Adams

Institute of Fluid Mechanics, Technical University Dresden, D-01062 Dresden, Germany
Justyna.Czerwinska@ism.mw.tu-dresden.de

1 Abstract

In this paper we present a new formulation of a Dissipative Particle Dynamics (DPD) model which is computationally less expensive than Voronoi-based DPD while preserving most of the advantages of Voronoi DPD over simple spherical-particle models. Aiming at fully three-dimensional flows an alternative to the straight-forward application of Voronoi DPD is desirable. The new model presented here can be derived from the Molecular Dynamics level by a coarse graining procedure (bottom-up approach) as well as from the continuum or macro-scale level conservation equations (top-down approach). In this paper the bottom-up derivation is presented.

2 Introduction

The Dissipative Particle Dynamics (DPD) method was first introduced by Hoogerbrugge and Koelman [HK92] with the intention of deriving a meso-scale fluid description on hydrodynamic time and space scales which are much larger than those on the Molecular Dynamics level. At the same time DPD avoids lattice artifacts of Lattice Gas Automata. Subsequent improvements lead to various formulations of the DPD method. Español and Warren [EW95] have proved that the original model of Hoogerbrugge and Koelman corresponds to a finite-difference approximation of a more general DPD formulation provided by a set of continuous-time Langevin equations. Bonet Avalos and Mackie [BAM97] defined meso-scopic entropy and temperature for a dissipative particle. In the formulation of Español and Revenga [ER03] particles have spherical size and fixed volume, although most of the DP properties are similar to the classical model. Most of these formulations have, however, deficiencies concerning, e.g., proper description of boundary conditions. As a remedy Flekkøy *et al.* [FCF00] proposed a DPD formulation based on a Voronoi tessellation of the flow domain. For the bottom-up derivation a Voronoi Dissipative Particle is considered as a cluster of MD particles. Serrano and Español [SE01] derived a similar model from the top-down approach by a discretization of the Navier-Stokes

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

equations. The Voronoi method solves most problems of previous versions of DPD. It is, however, computationally very expensive due to the need of tracking Voronoi-particle deformations [CA03]. DPD simulations based on Voronoi tessellation can be very inefficient for large 3D computations. The generation of a Voronoi diagram for a set of n points by the divide-and-conquer algorithm requires O(n log n) operations [SH75]. A more efficient sweepline algorithm with same operation count was presented by Fortune [For86]. Various Voronoi algorithms are reviewed by Okabe *et al.* [OBS92]. Most of the Voronoi algorithms work bestfor 2D applications but even then dynamic changes of the Voronoi structure cause considerable computational overhead.

The purpose of this paper is to derive an alternative DPD formulation based on a Voronoi lattice. The information about the tessellation is incorporated into the equations of particle motion without requiring the Voronoi connectivity explicitly. The implementation of this method is almost as straight forward as the spherical-particle DPD. The particles can be considered as set of a nearly-tight pack of spheres with different sizes. Also, boundary conditions can be formulated properly and a considerable increase of computational efficiency is achieved.

3 Meso-scopic description

For deriving the equations of meso-scale particle motion we follow essentially the procedure of [FCF00], with differences related to the shape of the particles.

3.1 Basic micro-scopic equations

A Dissipative Particle (DP) is defined as a cluster of Molecular Dynamics particles (MD). The volume of a DP may change due to an unrestricted motion of MD particles between DP. Following [FCF00] we define a sampling function

$$f_k(\mathbf{x}) = \frac{s(\mathbf{x}-\mathbf{r}_k)}{\sum_l s(\mathbf{x}-\mathbf{r}_l)} \ , \tag{1}$$

where $\mathbf{r}_k$, $\mathbf{r}_l$ are positions of DP centers and $\mathbf{x}$ is arbitrary position. The function $s(\mathbf{x})$ can be chosen as a Gaussian

$$s(\mathbf{x}) = \exp(-|\mathbf{x}|^2/a^2) \ . \tag{2}$$

The product of two Gaussian functions is again Gaussian. This property will be very helpful for deriving the equation of motions. With the sampling function mass, momentum and internal energy can be computed by summation over all Dissipative Particles as

$$M_k = \sum_i f_k(\mathbf{x}_i) m \ , \tag{3}$$

$$\mathbf{P}_k = \sum_i f_k(\mathbf{x}_i) m \mathbf{v}_i \ , \tag{4}$$

$$\frac{1}{2}M_k U_k^2 + E_k = \sum_i f_k(\mathbf{x}_i)\left(\frac{1}{2}mv_i^2 + \frac{1}{2}\sum_{j\neq i} V_{MD}(\mathbf{r}_{ij})\right) \equiv \sum_i f_k(\mathbf{x}_i)\epsilon_i \ , \quad (5)$$

where the index k refers to a DP particle and the index i to a MD particle. $V_{MD}(\mathbf{r}_{ij})$ is the potential energy of the MD particle pair ij, separated by distance $\mathbf{r}_{ij}$; ϵ_i is the total energy of a MD particle. A DP moves by its average velocity

$$\dot{\mathbf{r}}_k = \mathbf{U}_k = \mathbf{P}_k/M_k \ . \quad (6)$$

Since the sampling function is a partition of unity, mass, momentum and energy are conserved on the micro- and the meso-scale

$$\sum_k M_k = \sum_i m \ , \quad (7)$$

$$\sum_k \mathbf{P}_k = \sum_i m\mathbf{v}_i \ , \quad (8)$$

$$\sum_k E_k^{tot} = \sum_k \left(\frac{1}{2}M_k\mathbf{U}_k^2 + E_k\right) = \sum_i \epsilon_i \ . \quad (9)$$

Differentiation of equations (3), (4), (5) with respect to time gives

$$\frac{dM_k}{dt} = \sum_i \dot{f}_k(\mathbf{x}_i)m \ , \quad (10)$$

$$\frac{d\mathbf{P}_k}{dt} = \sum_i [\dot{f}_k(\mathbf{x}_i)m\mathbf{v}_i + f_k(\mathbf{x}_i)\mathbf{F}_i] \ , \quad (11)$$

$$\frac{dE_k^{tot}}{dt} = \sum_i [\dot{f}_k(\mathbf{x}_i)\epsilon_i + f_k(\mathbf{x}_i)\dot{\epsilon}_i] \ , \quad (12)$$

where the force on a MD particle i is $\mathbf{F}_i = m\dot{\mathbf{v}}_i$. Due to the fact that the sampling function is a Gaussian its derivative is

$$\dot{s}(\mathbf{x}_i) = -(2/a^2)\dot{\mathbf{x}}\mathbf{x}s(\mathbf{x}) \ . \quad (13)$$

Consequently

$$\dot{f}_k(\mathbf{x}_i) = f_{kl}(\mathbf{x}_i)(\mathbf{v}_i^P \cdot \mathbf{r}_{kl} + \mathbf{x}_i^P \cdot \mathbf{U}_{kl}) \ , \quad (14)$$

where the overlap function f_{kl} is given

$$f_{kl}(\mathbf{x}) = (2/a^2)f_k(\mathbf{x})f_l(\mathbf{x}) \ . \quad (15)$$

The distance between the centers of two neighboring DP k and l is $\mathbf{r}_{kl} = (\mathbf{r}_k - \mathbf{r}_l)$. The velocity difference is $\mathbf{U}_{kl} = (\mathbf{U}_k - \mathbf{U}_l)$. The center of the overlap region can be described in various ways. One possibility is

$$\mathbf{x}^P = \left(R_k + \frac{a}{1 + \frac{R_l}{R_k}}\right)\mathbf{c}_{kl} \ , \quad (16)$$

where $a = |\mathbf{r}_{kl} - (R_k + R_l)\,\mathbf{c}_{kl}|$, see also eq. (2), and R_k is a radius of k^{th} DP, $\mathbf{c}_{kl}$ is the unit vector parallel to $\mathbf{r}_{kl}$. The location of a MD particle with the respect to $\mathbf{x}^P$ is given by $\mathbf{x}_i^P = \mathbf{x}_i - \mathbf{x}^P$. Alternatively $\mathbf{x}^P$ can be obtained using the radical plane construction. The radical plane is the locus of points for which the tangential distance to the two spheres is equal,

$$\mathbf{x}^P = \frac{(\mathbf{r}_{kl})^2 + R_k^2 - R_l^2}{2(\mathbf{r}_{kl})^2}\mathbf{r}_{kl} \ . \tag{17}$$

The second method partitions space, however it is computationally more expensive than the first one.

A linear interpolation of function values (density, velocity, pressure) at point P is given by equation

$$h^P = h_l \frac{|\mathbf{x}^P - \mathbf{r}_k|}{|\mathbf{r}_{kl}|} + h_k \frac{|\mathbf{r}_l - \mathbf{x}^P|}{|\mathbf{r}_{kl}|} \ . \tag{18}$$

We defined the velocity $\mathbf{v}_i^P$ with respect to the velocity of P as $\mathbf{v}_i^P = \mathbf{v}_i - \mathbf{v}^P$. For the example for a Gaussian sampling function $f_k(\mathbf{x})$ is given by

$$f_k(\mathbf{x}) = \frac{1}{1 + (\exp(2\mathbf{x}^P \cdot \frac{\mathbf{r}_{kl}}{a^2}))^2} \ . \tag{19}$$

The corresponding overlap function is

$$f_{kl}(\mathbf{x}) = \frac{1}{2a^2}\cosh^{-2}\left[\mathbf{x}^P \cdot \frac{\mathbf{r}_{kl}}{a^2}\right] \ . \tag{20}$$

A sketch of sampling and overlap function is provided in Fig. 1. Note also that the overlap function for the next-to-nearest neighbor interaction is one order of magnitude smaller than that for the nearest-neighbor interaction. Therefore interactions are restricted to the nearest neighbors.

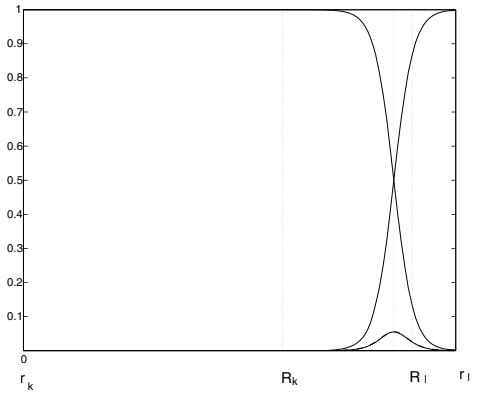


Fig. 1. 1D illustration of the sampling function and the overlap function for two DP at $\mathbf{r}_k$ and $\mathbf{r}_l$, respectively. R_k indicates the particle radius with origin at point $\mathbf{r}_k$; similarly for particle l.

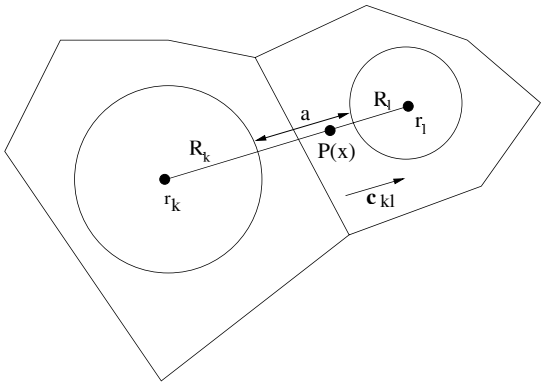


Fig. 2. Two DP $\mathbf{r}_k$ and $\mathbf{r}_l$ and their interaction region; $\mathbf{c}_{kl}$ is a unit vector parallel to $\mathbf{r}_{kl} = \mathbf{r}_k - \mathbf{r}_l$.

Using the above equations the system of conservation laws can be rewritten as

$$\frac{dM_k}{dt} = \sum_l \dot{M}_{kl} \ , \tag{21}$$

$$\dot{M}_{kl} = \sum_i f_{kl}(\mathbf{x}_i)\, m\, (\mathbf{v}_i^P \cdot \mathbf{r}_{kl} + \mathbf{x}_i^P \cdot \mathbf{U}_{kl}) \ . \tag{22}$$

The momentum equation can be written as:

$$\frac{d\mathbf{P}_k}{dt} = \sum_{l,i} f_{kl}(\mathbf{x}_i) m \mathbf{v}_i (\mathbf{v}_i^P \cdot \mathbf{r}_{kl} + \mathbf{x}_i^P \cdot \mathbf{U}_{kl}) + \sum_{l,i} f_k(\mathbf{x}_i)\mathbf{F}_i \ . \tag{23}$$

The force $\mathbf{F}_i$ is given by equation [FCF00]:

$$\sum_i f_k(\mathbf{x}_i)\mathbf{F}_i = M_k \mathbf{g} + \sum_{i,j} f_k(\mathbf{x}_i)\mathbf{F}_{ij} \ , \tag{24}$$

where $\mathbf{g}$ is the gravitational acceleration. Conservation laws on the microscopic level can be rewritten in the frame of reference of the overlap region. With Newton's third law the following simplification can be made

$$\begin{aligned} \sum_{i,j} f_k(\mathbf{x}_i)\mathbf{F}_{ij} = -\sum_{i,j} f_k(\mathbf{x}_i)\mathbf{F}_{ji} = -\sum_{i,j} f_k(\mathbf{x}_j + \Delta\mathbf{x}_{ij})\mathbf{F}_{ji} \approx \\ \approx -\sum_{i,j} f_k(\mathbf{x}_j)\mathbf{F}_{ji} - \sum_{i,j} [\Delta\mathbf{x}_{ij} \cdot \nabla f_k(\mathbf{x}_i)]\mathbf{F}_{ji} = \\ = -\frac{1}{2}\sum_{i,j} [\Delta\mathbf{x}_{ij} \cdot \nabla f_k(\mathbf{x}_i)]\mathbf{F}_{ji} = \sum_l \left(\sum_{i,j} \frac{1}{2} f_{kl}(\mathbf{x}_i)\mathbf{F}_{ij}\Delta\mathbf{x}_{ij} \right) \cdot \mathbf{r}_{kl} \ , \end{aligned} \tag{25}$$

where $\Delta\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$. The momentum equation can be rewritten in the form

$$\frac{d\mathbf{P}_k}{dt} = M_k\mathbf{g} + \sum_l \dot{M}_{kl}\mathbf{U}^P + \sum_{l,i} f_{kl}(\mathbf{x}_i)\mathbf{\Pi}_i^P \cdot \mathbf{r}_{kl} + \\ + \sum_{l,i} f_{kl}(\mathbf{x}_i) m\mathbf{v}_i^P \mathbf{x}_i^P \cdot \mathbf{U}_{kl} \ , \tag{26}$$

where $\mathbf{\Pi}_i$ is the general momentum-flux tensor

$$\mathbf{\Pi}_i = m\mathbf{v}_i\mathbf{v}_i + \frac{1}{2}\mathbf{F}_{ij}\Delta\mathbf{x}_{ij} \ . \tag{27}$$

The energy time rate of change for MD particle is

$$\dot{\epsilon}_i = m\mathbf{g}\cdot\mathbf{v}_i + \frac{1}{2}\sum_{j\neq i}\mathbf{F}_{ij}\cdot(\mathbf{v}_i + \mathbf{v}_j) \ . \tag{28}$$

The last term of the equation (12) is

$$\sum_i f_k(\mathbf{x}_i)\dot{\epsilon}_i = \mathbf{P}_k\cdot\mathbf{g} + \frac{1}{2}\sum_{i\neq j} f_k(\mathbf{x}_i)\mathbf{F}_{ij}\cdot(\mathbf{v}_i + \mathbf{v}_j) \ , \tag{29}$$

which can be re-written as

$$\sum_i f_k(\mathbf{x}_i)\dot{\epsilon}_i = \mathbf{P}_k\cdot\mathbf{g} + \sum_{l,i\neq j} f_{kl}(\mathbf{x}_i)\frac{1}{4}\mathbf{F}_{ij}\cdot(\mathbf{v}_i + \mathbf{v}_j)\Delta\mathbf{x}_{ij}\cdot\mathbf{r}_{kl} = \\ = \mathbf{P}_k\cdot\mathbf{g} + \sum_{l,i\neq j} f_{kl}(\mathbf{x}_i)\left(\frac{1}{4}\mathbf{F}_{ij}\cdot(\mathbf{v}_i^P + \mathbf{v}_j^P) + \frac{1}{2}\mathbf{F}_{ij}\cdot\mathbf{U}^P\right)\Delta\mathbf{x}_{ij}\cdot\mathbf{r}_{kl} \ . \tag{30}$$

The next-to-last term of equation (12) can be rewritten as

$$\sum_i \dot{f}_k(\mathbf{x}_i)\epsilon_i = \sum_{l,i} f_{kl}(\mathbf{x}_i)(\mathbf{v}_i^P\cdot\mathbf{r}_{kl} + \mathbf{x}_i^P\cdot\mathbf{U}_{kl})\epsilon_i \ . \tag{31}$$

The particle energy evaluated with the respect to point P is

$$\epsilon_i = \epsilon_i^P + m\mathbf{v}_i^P\cdot\mathbf{U}^P + \frac{1}{2}m(\mathbf{U}^P)^2 \ . \tag{32}$$

The above expressions can be combined to yield

$$\sum_i \dot{f}_k(\mathbf{x}_i)\epsilon_i = \sum_l \frac{1}{2}\dot{M}_{kl}(\mathbf{U}^P)^2 + \\ \sum_{l,i} f_{kl}(\mathbf{x}_i)\left(\epsilon_i^P\mathbf{v}_i^P + m\mathbf{v}_i^P\mathbf{v}_i^P\cdot\mathbf{U}^P\right)\cdot\mathbf{r}_{kl} + + \sum_{l,i} f_{kl}(\mathbf{x}_i)\epsilon_i\mathbf{x}_i^P\cdot\mathbf{U}_{kl} \ . \tag{33}$$

The total energy equation is

$$\frac{dE_k^{tot}}{dt} = \dot{E}_k^{tot} = \sum_{l,i} f_{kl}(\mathbf{x}_i)\big(\mathbf{J}_{\epsilon i}^P + \mathbf{\Pi}_i^P \cdot \mathbf{U}^P\big) \cdot \mathbf{r}_{kl} + \\ + M_k \mathbf{U}_k \cdot \mathbf{g} + \sum_l \frac{1}{2}\dot{M}_{kl}(\mathbf{U}^P)^2 + \\ + \sum_{l,i} f_{kl}(\mathbf{x}_i)\left[\epsilon_i^P + m\mathbf{v}_i^P \cdot \mathbf{U}^P\right]\mathbf{x}_i^P \cdot \mathbf{U}_{kl} \ . \tag{34}$$

The energy flux vector for the MD particle i is

$$\mathbf{J}_{\epsilon i} = \epsilon_i \mathbf{v}_i + \frac{1}{4}\sum_{i \neq j} \mathbf{F}_{ij} \cdot (\mathbf{v}_i + \mathbf{v}_j) \Delta \mathbf{x}_{ij} \ . \tag{35}$$

This set of equations represents conservation of mass, momentum and energy at the microscopic level. Next, meso-scale equations will be derived by coarse-graining the micro-scopic equations of motions.

3.2 Meso-scopic equations

The meso-scopic equations describe conservations laws for Dissipative Particles. In this section we will again follow the derivation provided by [FCF00] and point out mainly the differences to our method. A scale separation between MD particles and DP is assumed as

$$|\mathbf{x}_i - \mathbf{x}_j| \ll |\mathbf{r}_k - \mathbf{r}_l| \ , \tag{36}$$

$$|\mathbf{x}_i - \mathbf{x}_j| \ll (R_k - R_l) \ , \tag{37}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are positions of neighboring MD particles. The average velocity of a MD particle depends on the velocities of the two neighboring DP. Assuming linear interpolation, the average MD particle velocity is approximately

$$\langle \mathbf{v}_i^P \rangle = \frac{\mathbf{x}_i^P \cdot \mathbf{r}_{kl}}{r_{kl}^2}\mathbf{U}_{kl} \ . \tag{38}$$

Splitting the equations of motion into the average and fluctuating part requires some additional assumptions. The average of a function h over the overlap region can be approximated as:

$$\sum_i f_{kl}(\mathbf{x}_i)\, h(\mathbf{x_i}) \approx \int d\phi\, d\mathbf{x}\, |\mathbf{r}_{kl}|\, f_{kl}(\mathbf{x})\, h^P \approx 2\pi\, h^P\, \frac{|\mathbf{r}_{kl}|}{2a^2}\mathbf{x}^P \ , \tag{39}$$

where ϕ is the polar angle. The first moment of the function $h(\mathbf{x})$ with respect to $\mathbf{x}^P$ is

$$\sum_i f_{kl}(\mathbf{x}_i)\, \mathbf{x}_i^P\, h(\mathbf{x}_i) \approx \int d\phi\, d\mathbf{x}\, (\mathbf{x} - \mathbf{x}^P)\, |\mathbf{r}_{kl}|\, f_{kl}(\mathbf{x})\, h^P \approx \\ \approx 2\pi\, h^P\, \frac{|\mathbf{r}_{kl}|}{4a^2}(\mathbf{x}^P)^2 \cdot \mathbf{c}_{kl} \tag{40}$$

This approximation is obtained from Taylor series expansion in the vicinity of point $\mathbf{x}^P$. For illustration we show the 2D construction. The extension to 3D is straight forward. The averaging of the mass conservation equation leads to

$$\frac{dM_k}{dt} = \dot{M}_k = \sum_l \left(\langle \dot{M}_{kl} \rangle + \dot{\tilde{M}}_{kl} \right) , \tag{41}$$

where

$$\langle \dot{M}_{kl} \rangle = \sum_{l,i} f_{kl}(\mathbf{x}_i)\, m\, \langle \mathbf{x}_i^P \rangle \cdot \mathbf{U}_{kl} = \pi \, \frac{|\mathbf{r}_{kl}|}{a^2} \, (\mathbf{x}^P)^2 \, \rho_P \, \mathbf{c}_{kl} \cdot \mathbf{U}_{kl} \ , \tag{42}$$

and $\dot{\tilde{M}}_{kl} = \dot{M}_{kl} - \langle \dot{M}_{kl} \rangle$. The average momentum equation can be written in the form

$$\begin{aligned} \frac{d\mathbf{P}_k}{dt} = M_k \mathbf{g} + \sum_l \langle \dot{M}_{kl} \rangle \mathbf{U}^P + \sum_{l,i} f_{kl}(\mathbf{x}_i) \langle {\mathbf{\Pi}^{\mathbf{P}}}_i \rangle \cdot \mathbf{r}_{kl} + \\ + \sum_i f_{kl}(\mathbf{x}_i) m \langle \mathbf{v}_i^P \mathbf{x}_i^P \rangle \cdot \mathbf{U}_{kl} + \sum_l \tilde{\mathbf{F}}_{kl} \ , \end{aligned} \tag{43}$$

where the fluctuating force is

$$\begin{aligned} \tilde{\mathbf{F}}_{kl} = \sum_i f_{kl}(\mathbf{x}_i) [({\mathbf{\Pi}^{\mathbf{P}}}_i - \langle {\mathbf{\Pi}^{\mathbf{P}}}_i \rangle) \cdot \mathbf{r}_{kl} + \\ + m(\mathbf{v}_i^P \mathbf{x}_i^P - \langle \mathbf{v}_i^P \mathbf{x}_i^P \rangle) \cdot \mathbf{U}_{kl}] \end{aligned}$$

The momentum-flux tensor for a Newtonian fluid is

$$\rho \langle \mathbf{\Pi}_i \rangle = m \rho \mathbf{v}\mathbf{v} + \mathbf{I} p - \eta (\nabla \mathbf{v} + (\nabla \mathbf{v})^T) \ . \tag{44}$$

Assuming linear interpolation the advective term approximately vanishes in the frame of reference of the overlap region since $\mathbf{v}_i^P \approx 0$

$$\nabla \mathbf{v}^P + (\nabla \mathbf{v}^P)^T = \frac{1}{|\mathbf{r}_{kl}|} (\mathbf{c}_{kl} \cdot \mathbf{U}_{kl} + \mathbf{U}_{kl} \cdot \mathbf{c}_{kl}) \ . \tag{45}$$

Consequently the momentum equation can be written as:

$$\begin{aligned} \frac{d\mathbf{P}_k}{dt} = M_k \mathbf{g} + \sum_l \langle \dot{M}_{kl} \rangle \mathbf{U}^P - \\ - \sum_l \pi \frac{|\mathbf{r}_{kl}|}{a^2} \mathbf{x}^P \left(p^P \, \mathbf{c}_{kl} + \frac{\eta}{|\mathbf{r}_{kl}|} [\mathbf{U}_{kl} + (\mathbf{U}_{kl} \cdot \mathbf{c}_{kl}) \mathbf{c}_{kl}] \right) + \sum_l \tilde{\mathbf{F}}_{kl} . \end{aligned} \tag{46}$$

The average energy equation is

$$\frac{dE_k}{dt} = \dot{E}_k = \sum_{l,i} f_{kl}(\mathbf{x}_i)(\langle \mathbf{J}^P_{\epsilon i}\rangle + \langle \mathbf{\Pi}^P_i\rangle \cdot \mathbf{U}^P) \cdot \mathbf{r}_{kl} + \\ + \sum_{l,i} f_{kl}(\mathbf{x}_i)\langle \epsilon^P_i \mathbf{x}^P_i\rangle \cdot \mathbf{U}_{kl} + \sum_l \frac{1}{2}\langle \dot{M}_{kl}\rangle (\mathbf{U}^P)^2 + \sum_l \tilde{\mathbf{F}}_{kl} \cdot \mathbf{U}^P + \tilde{q}_{kl} \ , \tag{47}$$

where the fluctuating part is defined as

$$\tilde{q}_{kl} = \sum_i f_{kl}(\mathbf{x}_i)(\mathbf{J}^P_{\epsilon i} - \langle \mathbf{J}^P_{\epsilon i}\rangle) \cdot \mathbf{r}_{kl} + \frac{\dot{\tilde{M}}_{kl}}{2}(\mathbf{U}^P)^2 + \\ + \sum_i f_{kl}(\mathbf{x}_i) \times \left[(\epsilon^P_i \mathbf{x}^P_i - \langle \epsilon^P_i \mathbf{x}^P_i\rangle) + m\mathbf{U}^P \cdot \mathbf{v}^P_i \mathbf{x}^P_i \right] \cdot \mathbf{U}_{kl} \ . \tag{48}$$

The average energy flux is assumed to have the form

$$\rho \langle \mathbf{J}_\epsilon \rangle = \epsilon \mathbf{v} + \mathbf{c} \cdot \mathbf{v} - \lambda \nabla T \ . \tag{49}$$

With $\mathbf{v}^P_i \approx 0$ it is $\langle \mathbf{J}^{\mathbf{P}}_\epsilon \rangle = -\lambda \nabla \mathbf{T}$ and the average energy equation for the Newtonian fluid can be written as

$$\dot{E}_k = -\sum_l \pi \frac{|\mathbf{r}_{kl}|}{a^2} \mathbf{x}^P \lambda \frac{T_{kl}}{r_{kl}} + \sum_l \frac{1}{2}\langle \dot{M}_{kl}\rangle (\mathbf{U}^P)^2 + \\ + \sum_l \pi \frac{|\mathbf{r}_{kl}|}{a^2} \mathbf{x}^P \left(p^P \mathbf{c}_{kl} - \frac{\eta}{r_{kl}} \times [\mathbf{U}_{kl} + (\mathbf{U}_{kl} \cdot \mathbf{c}_{kl})\mathbf{c}_{kl}] \right) \cdot \mathbf{U}^P + \\ + \pi \frac{|\mathbf{r}_{kl}|}{2a^2} (\mathbf{x}^P)^2 \, \mathbf{c}_{kl} \cdot \mathbf{U}_{kl} \left(\frac{E_k}{V_k} + \frac{E_l}{V_l} \right) + \\ + \sum_l \tilde{\mathbf{F}}_{kl} \cdot \frac{\mathbf{U}_{kl}}{2} + \tilde{q}_{kl} \ . \tag{50}$$

4 Implementation issues

There are some important implementational issues connected with the above developed method. The first is the formulation of boundary conditions. Similarly to the Voronoi method boundary conditions can be imposed at the boundary itself, by specifying quantities on the dual mesh at the centers of the overlap function. The second concern is related to the efficiency of the algorithm. With the presented method there is still the requirement of finding nearest neighbors (similar to the Voronoi approach). These can be obtained, however, without actually computing the Voronoi tessellation, which is computationally very expensive, particularly in 3D. Moreover, the dual mesh needs twice less the amount of storage which would be required for representing the Voronoi edges. A third point is connected with the size of spherical meso-scale particles. An initial set of points (particle centers) can be a basis for a Voronoi lattice from which subsequently the non-overlapping spheres are obtained.

With the motion of the particles the size of the spheres needs to be adjusted. This can be accomplished in various ways. One option is to determine the location of the point P by keeping the initial ratio of distances between point P and the neighboring particle centers fixed, and the location of the point P can be estimated at later times. From this condition the radii of the spherical particle, can be constructed.

5 Conclusions

In this paper a new approach for a DPD method is presented. The method shares favorable properties with the Voronoi based DPD approach. It is, however, computationally more efficient. This improvement is obtained by replacing Voronoi cells by their corresponding spheres. The spherical particles interaction was modeled by a Gaussian sampling function, which allows to introduce a dual mesh. Similarly to [FCF00] this property was exploited to establish the meso-scopic equations of motion for Dissipative Particles. A computational validation of the method will be presented in a subsequent paper.

References

[BAM97] Bonet Avalos J., Mackie A. D.: Dissipative particle Dynamics with Energy Conservation. Europhys. Lett. **40**, 141–146 (1997)

[CA03] Czerwinska J., Adams N. A.: Numerical modeling of micro-channel flows by a DPD method, In: ASME FED'03, (2003)

[ER03] Español, P., Revenga M.: Smoothed dissipative particle dynamics. Phys. Rev. E **67**, 026705–1–12 (2003)

[EW95] Español P., Warren, P. B.: Statistical mechanics of dissipative particle dynamics. Europhys. Lett. **30**, 191–196 (1995)

[FCF00] Flekkøy E.G., Coveney P.V., de Fabritiis G.: Foundations of dissipative particle dynamics. Phys. Rev. E **62**, 2140–2157 (2000)

[For86] Fortune S.: A Sweepline algorithm for Voronoi Diagrams. In: 2nd Annu. ACM Symp. on Comp. Geom., 313–322 (1986)

[HK92] Hoogerbrugge P. J., Koelman J. M. V. A.: Simulating Microscopic Hydrodynamic Phenomena with Dissipative Particle Dynamics. Europhys. Lett., **19**, 155–160 (1992)

[OBS92] Okabe A., Boots B., Sugihara K.: Spatial Tessellation: Concepts and Applications of Voronoi Diagrams., John Wileyand Sons, Chichester, UK (1992)

[SE01] Serrano M., Español P.: Thermodynamically consistent mesoscopic fluid particle model. Phys. Rev. E **64**, 1–18 (2001)

[SH75] Shamos M. I., Hoey D.: Closest point problem. In: 16th Annu. IEEE Symp. Found. Comp. Sci., 151–162 (1975)

Population balance modeling of synthesis of nanoparticles in aerosol flame reactors

Stavros Tsantilis[1]

Particle Technology Laboratory,
Department of Mechanical and Process Engineering (D-MAVT)
ETH Zrich, Switzerland
stsanti@mail.student.ethz.ch

Summary. Population balance models typically employed in simulations of synthesis of nanoparticles in aerosol flame reactors are presented. The main focus of the present study is on sectional techniques for a variety of particle formation and growth mechanisms involving gas phase and surface chemical reactions, coagulation, and sintering. A moving sectional model is described and compared with other models including a detailed two dimensional sectional method. Conditions of applicability of the above mentioned numerical model are also investigated.

1 Introduction

The most important mechanisms of particle formation and growth are monomer inception (that is nucleation or gas phase chemical reaction), coagulation, sintering and surface growth or condensation. All these mechanisms, except sintering, are represented in the following general one-dimensional dynamic equation [Kum97]

$$\frac{\partial n(v,t)}{\partial t} + \frac{\partial (G(v)n(v,t))}{\partial v} = \frac{1}{2}\int_0^\infty n(v-v',t)n(v',t)q(v-v',v')dv' - n(v,t)\int_0^\infty n(v',t)q(v,v')dv' + S(v) \quad (1)$$

where v is the particle volume, $n(v,t)$ is the particle density function of particles between v and v + dv at time t, G(v) is the particle growth rate, $q(v,t)$ is the collision frequency between particles of volumes v and and S(v) is the particle nucleation rate. The following section focuses on some of the most characteristic modeling efforts on combinations of the above mechanisms. There are four basic types of models used in aerosol dynamics simulations: monodisperse, moment, discrete and sectional models. These are all numerical techniques arising from the lack of a general analytical solution of the dynamic equation (1). Monodisperse models neglect the spread of the particle size distribution by assuming particles or aggregates of equal size and

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

describe, reasonably well, integral properties of the size distribution such as average particle size and total number concentration [Kru93; Pan95; Tsa99; Tsa01]. They therefore constitute an attractive option in terms of simplicity and computational efficiency and can provide fast predictions for a number of aerosol applications.

Typically, moment models describe the evolution of the moments of the aerosol size distribution by approximating its shape, for instance as lognormal or self-preserving [Whi81], [Lan90]. Their accuracy strongly depends on the validity of the above mentioned approximations [Fre87]. More recent approaches by-pass a-priori assumptions on the shape of the particle size distribution [McG97], [Fre02]. Moment methods provide information again on integral properties as well as general characteristics of the particle size distribution such as the standard deviation, skewness and kurtosis [Fr85]. Discrete models provide the most accurate solutions to the detailed population balance equations by accounting for all particle sizes but have considerable computational requirements [Fre87]. Sectional models require less computational effort by discretizing the particle size distribution in a finite number of sections within which the size distribution density function remains constant. In addition sectional models can be combined with discrete schemes, especially when nucleation or birth of monomers by gas phase chemical reaction is to be taken into account [Rog97], [Nak01].

2 Theory - Moving Sectional Aerosol Dynamics Model

Of particular interest is the investigation of the fundamental mechanisms contributing to the evolution of aerosol-made particles, i.e. titania (TiO2), in one general model. Here, a sectional model encompassing all the major mechanisms, namely formation by gas phase and surface growth reactions followed by coagulation and sintering is described and compared with other models. Case studies will mainly include flame synthesis of TiO2 from titanium tetra-isopropoxide (TTIP) decomposition and silica SiO2 from SiCl4 oxidation. As demonstrated in earlier studies [Hou88]; [Kum96a];[Kum96b]; [Kum97]; [Tsa00]; [Jeo01], the considerable increase of computational power, along with the development of efficient models render sectional calculations an increasingly attractive simulation tool providing accurate and detailed information on the dynamics of particle size distributions. Here, the proposed sectional formulation accounts for simultaneous gas phase chemical reaction (or nucleation), surface growth, coagulation and sintering and constitutes an extension of the moving sectional model of Kumar and Ramkrishna [Kum97] to include a new discretization for nucleation by Chaoul [Cha00] and Spicer and coworkers [Spi02] as well as sintering and particle shape irregularity effects following the approach of Vemury and coworkers [Vem94]. The latter used the sectional model of Batterham and coworkers [Bat81] and proposed a sectional equation for the change in particle surface area concentration, A_i, by pure agglomeration (particle collisions without coalescence), by substituting the particle volume concentration in each size interval i with the corresponding particle surface area concentration. The total change of A_i by coagulation and sintering was then given as:

$$\frac{dA_i}{dt} = \frac{dA_i}{dt}\Big|_{\text{agg}} + \frac{dA_i}{dt}\Big|_{\sin t} \tag{2}$$

However, since the model of Batterham and coworkers [Bat81] contains systematic inaccuracies in the preservation of both total particle mass and numbers [Hou88], the moving sectional formulation of Kumar and Ramkrishna [Kum97] is more appropriate, especially when dealing with surface growth reaction modes that can produce (in the absence of a moving sectional technique) significant numerical diffusion errors. Therefore, the evolution (by coagulation) of the particle number concentration Ni (particles/g), in a given section i, is [Kum96a]; [Kum97]:

$$\frac{dN_i}{dt}|_{agg} = \sum_{x_{i-1}\le x_j+x_k\le x_i}^{i\ge j\ge k} \left(1-\frac{1}{2}\delta_{j,k}\right)\beta_{j,k}N_jN_k\left(\frac{x_{i-1}-v}{x_{i-1}-x_i}\right)\rho_g$$

$$+\sum_{x_i\le x_j+x_k\le x_{i+1}}^{i\ge j\ge k} \left(1-\frac{1}{2}\delta_{j,k}\right)\beta_{j,k}N_jN_k\left(\frac{x_{i+1}-v}{x_{i+1}-x_i}\right)\rho_g - N_i\sum_{k=1}^{M}\beta_{i,k}N_k\rho_g \tag{3}$$

where M is the total number of sections (bins), $d_{j,k}$ is the so called Dirac delta function (equal to 1 for j = k, and equal to 0 for j $\neq$ k), x_i (cm^3) is the pivot volume in section i with boundaries v_i and v_{i+1}, $v = x_j + x_k$ and $b_{j,k}$ (cm^3/s) is the Brownian collision frequency function between particles in pivots j and k, spanning over the free-molecular, transition, and continuum regimes [Fuc64]; [Sei86]. Equation 3 is a simplification of the general equation shown in Kumar and Ramkrishna [Kum96a]; [Kum97], representing here the case of preservation of mass and numbers. Similarly to Vemury and coworkers [Vem94], the area concentration in each section is:

$$\frac{dA_i}{dt}|_{agg} = \sum_{x_{i-1}\le x_j+x_k\le x_{i+1}}^{i\ge j\ge k} \left(1-\frac{1}{2}\delta_{j,k}\right)\beta_{j,k}\left(N_jA_k+N_kA_j\right)\eta'\cdot\rho_g$$

$$-A_i\sum_{k=1}^{M}\beta_{i,k}N_k\rho_g \tag{4}$$

where now (for preservation of mass and numbers) the volume (interpolation) fractions become:

$$\eta' = \begin{cases} \frac{\frac{x_{i+1}}{(x_j+x_k)}-1}{\frac{x_{i+1}}{x_i}-1} & x_i \le x_j+x_k \le x_{i+1} \\ \frac{\frac{x_{i-1}}{(x_j+x_k)}-1}{\frac{x_{i-1}}{x_i}-1} & x_{i-1} \le x_j+x_k \le x_i \end{cases} \tag{5}$$

When sintering, gas and surface chemical reactions are also considered, Equations 3 and 4 become:

$$\frac{dN_i}{dt} = k_gC\eta_i + \frac{dN_i}{dt}\Big|_{\text{agg}} \tag{6}$$

$$\eta_i = \begin{cases} 1 & v_m \in [v_i, v_{i+1}] \\ 0 & v_m \notin [v_i, v_{i+1}] \end{cases} \tag{7}$$

$$\frac{dA_i}{dt} = k_g C \eta_i \alpha_m + \sum_{x_{i-1} \le x_j + x_k \le x_{i+1}}^{i \ge j \ge k} \left(1 - \frac{1}{2}\delta_{j,k}\right) \beta_{j,k} \left(N_j A_k + N_k A_j\right) \eta' \cdot \rho_g$$
$$- A_i \sum_{k=1}^{M} \beta_{i,k} N_k \rho_g - \frac{1}{\tau_{s,i}} \left(A_i - N_i \alpha_{i,s}\right) + 4\pi N_i n_{p,i} k_s d_{p,i} C v_m \rho_g \quad (8)$$

where vm (cm^3) is the volume of a (i.e., TiO2) monomer equivalent sphere (3.32 10-23 cm3, corresponding to a volume equivalent diameter, dm, of approximately 0.4 nm), am (cm^2) is the corresponding TiO2 monomer area, ai,s (cm2) is the area of a fully fused TiO2 particle of volume x_i, $d_{p,i}$ $(cm) = (6N_i x_i)/A_i$ is the primary particle diameter in an aggregate of volume x_i [Kru93]; [Tsa00], $n_{p,i} = A_i^3/(36\pi\, x_i^2\, N_i^3)$ is the number of primary particles within that aggregate [Kru93]; [Tsa00] and τ_s,i $(s) = 7.4\, 10^8\, d_{p,i}^4\, T\, \exp(31000/T)$ is the characteristic sintering time (for titania by Kobata and coworkers [Kob91]). More specifically, the first RHS term in Equation 5 (for i = 1) accounts for the birth of particles of monomer size by gas phase chemical reaction [Tsa02] as titania particles in aerosol processes are typically formed in the absence of a nucleation barrier [Xio91]. Likewise, the first RHS term in Equation 6 accounts for the increase of particle area concentration by gas phase chemical reaction (nucleation), the second and third RHS terms account respectively for changes in aggregate area concentration by birth and death of particles in section i by coagulation, the fourth RHS term accounts for the effect of sintering [Koc90]; [Xio3]; [Vem94]; [Pan95], while the last RHS term in Equation 6 represents the increase in aggregate area by surface growth [Tsa02].

In order to avoid numerical diffusion when surface reactions take place, a moving sectional discretization for surface growth is also adopted [Kum97]:

$$\left.\frac{dV_i}{dt}\right|_{\text{surf}} = N_i \left.\frac{dx_i}{dt}\right|_{\text{surf}} = k_s C A_i\, v_m \rho_g \Rightarrow$$
$$\left.\frac{dx_i}{dt}\right|_{\text{surf}} = \frac{k_s C A_i v_m \rho_g}{N_i} = k_s C n_{p,i} \pi d_{p,i}^2 v_m \rho_g \quad (9)$$

In addition, for the effect of nucleation on the evolution of pivots, dx_i/dt, that preserves both mass and numbers [Cha00]; [Spi02] is:

$$\frac{dx_i}{dt} = \left.\frac{dx_i}{dt}\right|_{\text{nuc}} + \left.\frac{dx_i}{dt}\right|_{\text{surf}} \quad (10)$$

The system of differential equations to be solved for simultaneous nucleation, surface growth, coagulation and sintering include Equations 5, 6 and 8 (along with the reactions of the precursor) while the grid regeneration principles remain virtually the same as in Kumar and Ramkrishna [Kum97]. In general, the goal is to regenerate the grid preserving total mass, numbers and area without decreasing the accuracy of the calculations and maintain a rather geometric resolution, as this guarantees low computational demands for a wide range of particle sizes; thus for instance, pivot x_i is eliminated when the ratio of the adjacent pivots x_{i+1}/x_{i-1} is smaller than a critical value $r_{\text{critical}} = s^{1.1}$ where s is the initial spacing factor x_{i+1}/x_i.

3 Discussion

One-dimensional (1D) sectional models are usually developed in particle volume coordinates, while two-dimensional models are based on particle volume and area coordinates and are typically used as standard models for comparison with other schemes accounting for coagulation and sintering. Previous comparisons of the predictions of the present (1D) sectional model with those of the detailed but cumbersome (2D) scheme of Xiong and Pratsinis [Xio3] showed good agreement [Tsa02]. Here, comparisons in terms of detailed gas-stream primary and aggregate particle size distributions are also discussed. Figures 1 and 2 show predictions (solid lines) of aggregate particle size distributions from the 2D sectional calculations of Jeong and Choi [Jeo01] (Table 5) using the model of Xiong and Pratsinis [Xio3] for isothermal coagulation and sintering of SiO2, in the absence (Figure 1) or presence (Figure 2) of gas phase SiCl4 decomposition.

When the present model is used without any modifications (i.e., on the collision kernel) it predicts narrower aggregate particle size distributions especially in the absence of chemical reaction (Figures 1 and 2, dash double-dot lines). However, the collision kernel in Jeong and Choi [Jeo01] depends on a surface fractal dimension and an aggregate surface area accessibility factor [Xio3] rather than on a single mass fractal dimension [Tsa00]; [Tsa02]. Hence, the agreement with the calculations of Jeong and Choi [Jeo01] improves considerably when exactly the same coagulation rate as in Jeong and Choi [Jeo01] is employed (Figures 1 and 2, dash-dot lines). The difference between the two models however is more pronounced at small particle diameters between $4\,10^{-4}$mm and 10^{-3} mm (Figure 2). Nakaso and coworkers [Nak01] developed a discrete sectional version of the 2D sectional model of Xiong and Pratsinis [Xio3] and compared simulation results on coagulation and sintering of TiO2 (from TTIP or TiCl4 decomposition) with experimental data in a tubular

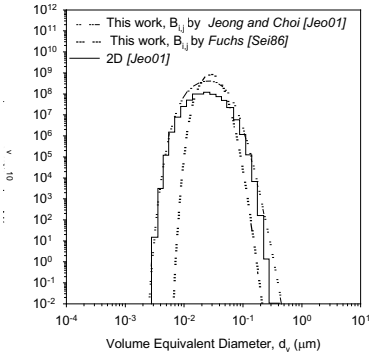


Fig. 1. Aggregate particle size distributions from the 2D sectional (solid line) calculations of Jeong and Choi [Jeo01] and the present 1D model (using two different collision kernels) for isothermal (T = 1800 K) coagulation and sintering of SiO2, in the absence of gas phase SiCl4 decomposition and electric fields.

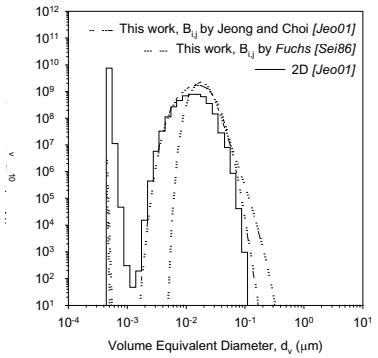


Fig. 2. Aggregate particle size distributions from the 2D sectional (solid line) calculations of Jeong and Choi [Jeo01] and the present 1D model (using two different collision kernels) for isothermal (T = 1750 K) coagulation and sintering of SiO2, in the presence of gas phase SiCl4 decomposition and no electric fields.

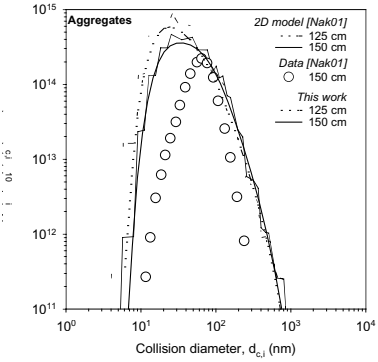


Fig. 3. Experimental data (open circles) of aggregate particle size distributions regarding synthesis of TiO2 from TTIP gas phase chemical reaction, coagulation and sintering, at a maximum reaction temperature of 1473 K and two different locations inside an aerosol furnace reactor [Nak01; Figure 9-2a] along with predictions from the 2D discrete-sectional model of Nakaso and coworkers [Nak01] (thin lines), and the present sectional model (thick lines).

hot wall reactor. Figure 3 shows their model predictions (thin lines) and experimental data (open circles) for aggregate particle size distributions regarding synthesis of TiO2 from TTIP at a maximum reaction temperature of 1473 K and two different locations inside the aerosol furnace reactor [Nak01] (Figure 9-2a). The predictions of the present sectional model, for the limiting case of precursor gas phase chemical reaction, coagulation and sintering and without any further modifications (thick lines), are in good agreement with those of Nakaso and coworkers [Nak01]. Nevertheless, the corresponding primary particle size distributions (Figure 4, thick lines) are narrower not only with respect to their experimental counterparts (Figure 4, filled cir-

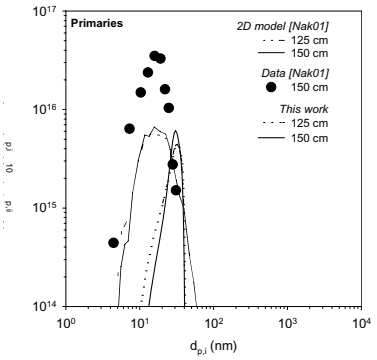


Fig. 4. Experimental data (filled circles) of primary particle size distributions regarding synthesis of TiO2 from TTIP gas phase chemical reaction, coagulation and sintering, at a maximum reaction temperature of 1473 K and two different locations inside an aerosol furnace reactor [Nak01; Figure 9-2a] along with predictions from the 2D discrete-sectional model of Nakaso and coworkers [Nak01] (thin lines), and the present sectional model (thick lines).

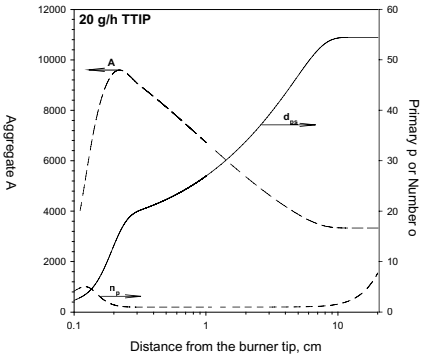


Fig. 5. Aggregate area concentration (broken line), primary particle Sauter mean diameter (solid line) and average number of primary particles per aggregate (dash and dot line) as predicted by the present sectional model, regarding titania particle growth by simultaneous gas phase and surface TTIP reactions followed by coagulation and sintering in a premixed methane flame [Kam03] for an inlet TTIP flowrate of 20 g/h).

cles) but also with respect to the predictions of Nakaso and coworkers [Nak01] (Figure 4, thin lines). This trend is consistent with earlier observations [Tsa02] showing that the narrowing of primary particle size distributions can be artificially enhanced under the assumption of a single particle area per volume size interval (one dimensional sectional model), especially for conditions of pure agglomeration ($t_{s,i} \rightarrow \infty$, A = constant, d_{ps} = constant, and $n_p > 1$) of initially polydisperse particles. For instance, this is clearly demonstrated in Figures 5 and 6 regarding titania particle growth by simultaneous gas phase and surface TTIP reactions followed by coagula-

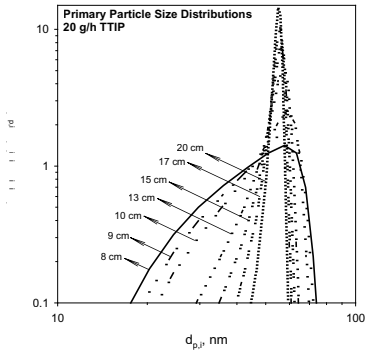


Fig. 6. Primary particle size distributions as predicted by the present sectional model for the same conditions as in Figure 5.

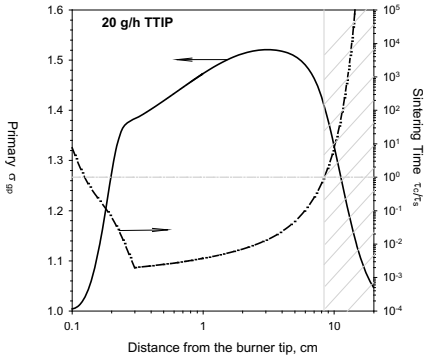


Fig. 7. Gas-stream primary particle geometric standard deviation (sgp, solid line) along with the ratio (dash and dot line) of the characteristic sintering time (ts dps) over the characteristic collision time, tc = 2/(bNrg), as predicted by the present sectional model for the same conditions as in Figure 5. The grey line pattern represents the region of soft agglomeration where A = constant and ts / tc ¿ 1 (grey dash-dot line).

tion and sintering in a premixed methane flame. Although pure agglomeration starts to evolve at approximately 8 cm above the burner tip (Figure 5), the corresponding primary particle size distribution continues to narrow more, even beyond x = 8 cm (Figure 6). This stems from the assignment of a (single) average primary particle diameter ($d_{p,i} = 6x_i/a_i$) in each volume size interval regardless of the initial polydispersity of the constituent primary particles. Figure 7 shows the evolution of the gas-stream primary particle geometric standard deviation (σ_{gp}, solid line) along with the ratio (dash and dot line) of the characteristic sintering time (ts dps) over the characteristic collision time, $\tau_c = 2/(bNr_g)$ [Tsa01], for the same conditions as in Figure 5. The above ratio is used here as a mathematical tool of tracking down the onset of pure agglomeration ($\tau_s/\tau_c > 1$ and A = constant) and consequently the re-

gion of applicability of the present model with respect to the prediction of primary particle size distributions or their corresponding geometric standard deviations. For the case of 20 g/h TTIP, pure agglomeration of polydisperse particles manifests itself at approximately 8 cm above the burner (Figures 5 and 7).

For $x > \approx 8cm$, particles collide but do not coalesce at all (thus, giving rise to the formation of soft agglomerates), leaving the primary particle size distribution unchanged. Therefore, the correct evolution of σ_{gp} in the region between 8 and 20 cm above the burner tip should be a straight line corresponding to a value of 1.4. More accurate predictions of primary particle size distributions can be achieved by expanding the current sectional model of Tsantilis and coworkers [Tsa02] from one to two dimensions, thus covering both particle volume and surface area coordinates and allowing for interpolations between two-dimensional sections. This approach (currently under investigation) provides a more simplified and computationally efficient version of the cumbersome two-dimensional sectional population balance model of Xiong and Pratsinis [Xio3] as it incorporates the so-called mean value theorem on frequency [Hou88]; [Kum96a]; [Kum96b] rather than the mean value theorem on number density [Gel80], therefore substituting the calculation of integrals in each section with a corresponding average value. In addition, it constitutes a more general particle dynamics model as it encompasses surface growth kinetics by incorporating a moving grid technique [Kum97] eliminating at the same time numerical diffusion problems stemming from fixed area sections.

4 Conclusions

The goal of this study was to present computationally efficient quantitative tools accounting for product powder characteristics, mainly associated with average particle size, polydispersity (ranging from molecular sizes to particle size scales larger than 100 nanometers) and degree of aggregation. A sectional representation of the general particle population balance equation was described accounting for a variety of concurrent formation and growth mechanisms including gas phase (nucleation) and surface chemical reactions, coagulation and sintering. Although the agreement with experimental data as well as other detailed sectional models was fairly good, the performance of the current model could be further improved to circumvent the inherent discrepancies arising from the simplified association of one aggregate particle volume with a single aggregate particle surface area. More accurate predictions of primary particle size distributions could be achieved by expanding the current sectional model from one to two dimensions, thus covering both particle volume and surface area coordinates and allowing for interpolations between two-dimensional sections.

5 Acknowledgements

Support by the Swiss National Science Foundation (Grant No. 2100-055469.98/1) and the Swiss Commission for Technology and Innovation (KTI) (TopNano21, Grant No- 5978.2TNS), as well as Prof. Sotiris E. Pratsinis for his thoughtful comments are gratefully acknowledged.

References

[Bat81] Batterham, R. J., Hall, J. S., and Barton, G. (1981) Pelletizing Kinetics and Simulation of Full Scale Balling Circuits. Proceedings, 3rd International Symposium on Agglomeration, Nurnberg, Federal Republic of Germany, A136.

[Cha00] Chaoul, O. (2000) Aerosol Process Simulation by Fixed and Moving Sectional Techniques. M. Sc. Thesis, Department of Chemical Engineering, University of Cincinnati, Cincinnati, OH 45221-0171, USA.

[Fr85] Frenklach, M. (1985) Dynamics of Discrete Distribution for Smoluchowski Coagulation Model. J. Colloid Interface Sci., 108, 237-242.

[Fre87] Frenklach, M., and Harris, S. J. (1987) Aerosol Dynamics Modeling Using the Method of Moments. J. Colloid Interface Sci., 118, 252-261.

[Fre02] Frenklach, M. (2002) Method of Moments with Interpolative Closure. Chem. Eng. Sci., 57, 2229-2239.

[Fuc64] Fuchs, N. A.(1964) The Mechanics of Aerosols, Pergamon Press, Elmsford, NY.

[Gel80] Gelbrand, F., Tambour, Y., and Seinfeld, J. H. (1980) Sectional Representations for Simulating Aerosol Dynamics. J. Colloid Interface Sci., 76, 541-556.

[Hou88] Hounslow, M. J., Ryall, R. L., and Marshall, V. R. (1988) A Discretized Population Balance for Nucleation, Growth and Aggregation. AIChE J., 34, 1821-1832.

[Jeo01] Jeong, J. I., and Choi, M. (2001) A Sectional Model for the Analysis of Growth of Non-Spherical Particles undergoing Coagulation and Coalescence. J. Aerosol Sci., 32, 565-582.

[Kam03] Kammler, H. K., Jossen, R., Morrison, P. W., Pratsinis , S. E., and Beaucage, G. (2003) The Effect of External Electric Fields during Flame Synthesis of Titania. Powder Technol., in press.

[Kob91] Kobata, A., Kusakabe, K., and Morooka, S. (1991) Growth and Transformation of TiO2 Crystallites in Aerosol Reactor. AIChE J., 37, 347-359.

[Koc90] Koch, W., and Friedlander, S. K. (1990) The Effect of Particle Coalescence on the Surface Area of a Coagulating Aerosol. J. Colloid Interface Sci., 140, 419-427.

[Kru93] Kruis, F. E., Kusters, K. A., and Pratsinis, S. E. (1993) A Simple Model for the Evolution of the Characteristics of Aggregate Particles Undergoing Coagulation and Sintering. J. Aerosol Sci. Tech., 19, 514-526.

[Kum96a] Kumar, S., and Ramkrishna, D. (1996a) On the Solution of Population Balance Equations by Discretization-I. A Fixed Pivot Technique. Chemical Eng. Sci., 51, 1311-1332.

[Kum96b] Kumar, S., and Ramkrishna, D. (1996b) On the Solution of Population Balance Equations by Discretization-II. A Moving Pivot Technique. Chemical Eng. Sci., 51, 1333-1342.

[Kum97] Kumar, S., and Ramkrishna, D. (1997) On the Solution of Population Balance Equations by Discretization-II. Nucleation, Growth and Aggregation of Particles. Chemical Eng. Sci., 52, 4659-4679.

[Lan90] Landgrebe, J., and Pratsinis, S. E. (1990) A Discrete-Sectional Model for Particulate Production by Gas-Phase Chemical Reaction and Aerosol Coagulation in the Free-Molecular Regime," J. Colloid Interface Sci., 139, 63-86.

[Nak01] Nakaso, K., Fujimoto, T., Seto, T., Shimada, M., Okuyama K., and Lunden, M. (2001) Size Distribution Change of Titania Nano-Particle Agglomerates Generated by Gas Phase Reaction, Agglomeration, and Sintering. Aerosol Sci. Technol., 35, 929-947.

[McG97] McGraw, R. (1997) Description of Aerosol Dynamics by the Quadrature Method of Moments. Aerosol Sci. Technol., 27, 255-265.

[Pan95] Panda, S., and Pratsinis, S. E. (1995) Modeling the Synthesis of Aluminum Particles by Evaporation-Condensation in an Aerosol Flow Reactor. Nanostructured Mater., 5, 755-767.

[Rog97] Rogak, S. N. (1997) Modeling Small Cluster Deposition on the Primary Particles of Aerosol Agglomerates. Aerosol Sci. Technol., 26, 127-140.

[Sei86] Seinfeld, J. H. (1986) Atmospheric Chemistry and Physics of Air Pollution, John Wiley & Sons, New York.

[Spi02] Spicer, P. T., Chaoul, O., Tsantilis, S., and Pratsinis, S. E. (2002) Titania Formation by TiCl4 Gas Phase Oxidation, Surface Growth and Coagulation. J. Aerosol Sci., 33, 17-34.

[Tsa99] Tsantilis, S., Pratsinis, S. E., and Haas, V. (1999) Simulation of Synthesis of Pd Nanoparticles in a Jet Aerosol Flow Condenser. J. Aerosol Sci., 53, 785-803.

[Tsa00] Tsantilis, S., and Pratsinis, S. E. (2000) Evolution of Primary and Aggregate Particle-Size Distributions by Coagulation and Sintering. AIChE J., 46, 407-415.

[Tsa01] Tsantilis, S., Briesen, H., and Pratsinis, S. E. (2001) Sintering Time for Silica Particle Growth. Aerosol Sci. Technol., 34, 237-246.

[Tsa02] Tsantilis, S., Kammler, H. K., and Pratsinis, S. E. (2002) Population Balance Modeling of Synthesis of Titania Nanoparticles. Chem. Eng. Sci., 57, 2139-2156.

[Vem94] Vemury, S., Kusters, K. A., and Pratsinis, S. E. (1994) Modeling of Coagulation and Sintering of Particles. Proceedings of the First International Particle Technology Forum, 2, 350-355.

[Whi81] Whitby, K. T. (1981) Determination of Aerosol Growth Rates in the Atmosphere Using Lumped Mode Aerosol Dynamics. J. Aerosol Sci., 12, 173-178.

[Xio91] Xiong, Y., and Pratsinis, S. E. (1991) Gas Phase Production of Particles in Reactive Turbulent Flows. J. Aerosol Sci. 22, 637-655.

[Xio3] Xiong, Y., and Pratsinis, S. E. (1993) Formation of Agglomerate Particles by Coagulation and Sintering-Part I. A Two Dimensional Solution of the Population Balance Equation. J. Aerosol Sci., 24, 283-300.

Modelling gene expression using stochastic simulation

Lars Kuepfer and Uwe Sauer

Institute of Biotechnology, ETH Zürich, HPT 8093 Zürich Switzerland
kuepfer@biotech.biol.ethz.ch

Abstract

Deterministic simulation of biological processes represents only a loose description of the actual intracellular mechanisms due to the small number of many molecular species involved in the regulatory circuits. Mesoscopic modelling that considers the systemic key species as integer numbers on a statistical basis was used in the present case study to solve a two gene sample problem in a eucaryotic cell. The results obtained with Gillespie's stochastic simulation algorithm were compared to the deterministic integration.

1 Introduction

The modelling of biological events on a molecular level is extremely important for comprehensive analysis and eventually understanding of the underlying mechanisms [SC03, SH03]. Simulation on such an explicit degree of presentation allows quantitative insights which are essential for testing of hypotheses or metabolic reprogramming and optimization [JB91, VK99, JB93]. Furthermore these models are extremely robust when the environmental or systemic conditions of the simulations are changed, for example when genes are deleted or overexpressed [BO86].

The model considered represents a small sample problem, encompassing the transcription and translation of two single genes, *TF1* and *TF2*, respectively (see Fig. 1). The species involved in the present system are highly coupled, as both protein products (Tf1, Tf2) are transcription factors themselves and they mutually affect the degree of transcription. Tf1 represents an activator that enhances gene expression, whereas Tf2 represents a repressor that acts as an inhibitor by competitively occupying possible binding sites on the genes. Both proteins are additionally self-regulating, as they associate with their own binding sites on the gene. The whole system is triggered by a master transcription factor Tfm, that acts as an activator on both genes *TF1* and *TF2*. A eukaryotic cell is modelled, where the genes are assembled in the nucleus. Transcription takes place in this compartment, which results in

T. J. Barth et al. (eds.), *Multiscale Modelling and Simulation*

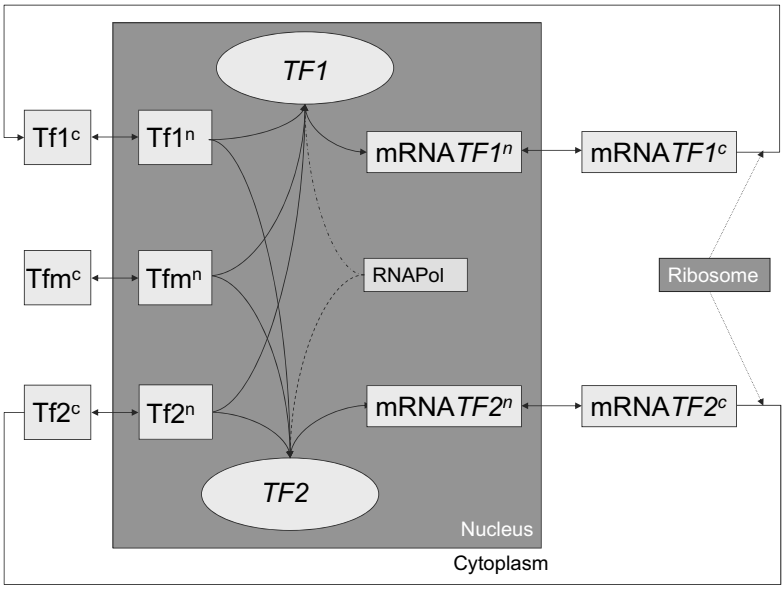


Fig. 1. A two gene network including compartmentalization in nucleus (^{n}) and cytoplasm (^{c}). *TF1* and *TF2* are genes, Tf1 and Tf2 the corresponding proteins. RNAPol and Ribosomes are molecules and cell organelles, respectively, that are present in excess.

the release of the specific mRNA. These molecules then migrate to the cytoplasm, where they are translated into proteins at the ribosomes. Compartmentalization is one way of genetic regulation in eukaryotic cells. Here, transcription factors of genes whose product is not needed are retained in the cytoplasm and only migrate into the nucleus upon demand [CO02].

2 Modelling gene expression using elementary reactions

To find an accurate representation of biological mechanisms, elementary reactions describing the governing steps of gene expression on a molecular level were formulated (see Tab. 1). Each step generally includes association and possible dissociation of the two reaction partners and a final irreversible catalytic step. Transcription, translation and decay of Tf1 and Tf2 was taken into account while the overall concentration of Tfm was assumed to be constant. All reactions were supposed to be of first order. Note, that there is only a single copy of each gene per cell, around 10 molecules of mRNA and 10-200 molecules of possible transcription factors, respectively. RNAPolymerase (RNAPol) and ribosomes are intracellular molecules and organelles that are present in excess.

The mass balances for each of the species lead to a system of coupled, non-linear ODEs (see Eqn.3-9). If one deals with molecular concentrations, the system pre-

Table 1. Modelling gene expression via elementary reactions

TRANSCRIPTION		
$[Tf] + [GENE]$	$\underset{K_{D1}}{\overset{K_{A1}}{\rightleftarrows}}$	$[Tf \cdot GENE]$
$[Tf \cdot GENE] + [RNAPol]$	$\underset{K_{D2}}{\overset{K_{A2}}{\rightleftarrows}}$	$[Tf \cdot GENE \cdot RNAPol]$
$[Tf \cdot GENE \cdot RNAPol]$	$\overset{K_{tc}}{\longrightarrow}$	$[Tf \cdot GENE] + [RNAPol] + [mRNA]$
TRANSLATION		
$[mRNA] + [Ribosome]$	$\underset{K_{D3}}{\overset{K_{A3}}{\rightleftarrows}}$	$[mRNA \cdot Ribosome]$
$[mRNA \cdot Ribosome]$	$\overset{K_{tl}}{\longrightarrow}$	$[mRNA] + [Protein] + [mRNA]$
DECAY		
$[mRNA]$	$\overset{K_{d,m}}{\longrightarrow}$	$\oslash$
$[Protein]$	$\overset{K_{d,P}}{\longrightarrow}$	$\oslash$

sentation is deterministic. For parametrization, gross estimations of the parameters within the expected order of magnitude were chosen (see Tab. 2.).

Altogether 26 species are to be considered: 14 single molecules, 8 dimeric and 4 trimeric complexes. Note, that due to the compartmentalization of the cell, unbound mRNAs and proteins occur both in the nucleus and in the cytoplasm. If one assumes a constant volume and a homogenous spatial distribution of the molecular species within the cell, the state variables represent average concentrations [M]. The size of the model could be reduced by considering the binding equations to be in rapid equilibrium, e.g.

$$Keq_{1,i,j} = \frac{k_{A1,i,j}[TFi][Tfj^n]}{k_{D1,i,j}[TFi \cdot Tfj^n]}. \tag{1}$$

This formulation leads to a differential algebraic set of equations (DAE) which, despite an index of 1, would be much harder to solve than the original system of ODEs. Furthermore, DAEs would mean additional manual work each time systemic or environmental conditions of the simulations are changed and therefore weaken the manageability and robustness of the model (see introduction).

The initial conditions of the species concentrations were chosen within an order of magnitude that is to be expected in a real cell. One can calculate a deterministic solution based on absolute molecule numbers n_i when the molar concentrations c_i

are converted by means of the Avogadro constant N_A and the corresponding cellular volume V (see Tab. 2.):

$$n_i = V \cdot N_A \cdot c_i. \tag{2}$$

$$\frac{d[\mathit{TFi}]}{dt} = \sum_{j=1}^{3}\Big(-k_{A1,i,j}[\mathit{TFi}][Tfj^n] + k_{D1,i,j}[\mathit{TFi} \cdot Tfj^n]\Big) \tag{3}$$

$$\begin{aligned}\frac{d[\mathit{TFi} \cdot Tfj^n]}{dt} = &\sum_{j=1}^{3}\Big(k_{A1,i,j}[\mathit{TFi}][Tfj^n] - k_{D1,i,j}[\mathit{TFi} \cdot Tfj^n]\Big) - ... \\ &\Big(\sum_{a=1}^{2} k_{A2,i,a}[\mathit{TFi} \cdot Tfa^n][RNAPol] - ... \\ &k_{D2,i,a}[\mathit{TFi} \cdot Tfa^n \cdot RNAPol]\Big)\end{aligned} \tag{4}$$

$$\begin{aligned}\frac{d[\mathit{TFi} \cdot Tfj^n \cdot RNAPol]}{dt} = &\Big(\sum_{a=1}^{2} k_{A2,i,a}[\mathit{TFi} \cdot Tfa^n][RNAPol] - ... \\ &k_{D2,i,a}[\mathit{TFi} \cdot Tfa^n \cdot RNAPol]\Big) - ... \\ &\sum_{a=1}^{2} k_{tc,i}[\mathit{TFi} \cdot Tfa^n \cdot RNApol]\end{aligned} \tag{5}$$

$$\begin{aligned}\frac{d[mRNA\mathit{TFi}^n]}{dt} = &\sum_{a=1}^{2} k_{tc,i}[tfi \cdot Tfa^n \cdot RNApol] - ... \\ &+k^{+}_{nt,m}[mRNA\mathit{TFi}^c] - k^{-}_{nt,m}[mRNA\mathit{TFi}^n] - ... \\ &k_{d,m}[mRNA\mathit{TFi}^n]\end{aligned} \tag{6}$$

$$\begin{aligned}\frac{d[mRNA\mathit{TFi}^c]}{dt} = &k^{-}_{nt,m}[mRNA\mathit{TFi}^n] - k^{+}_{nt,m}[mRNA\mathit{TFi}^c] - ... \\ &k_{d,m}[mRNA\mathit{TFi}^-] - k_{A3,i}[mRNA\mathit{TFi}^n][Ribosome] + ... \\ &k_{D3,i}[mRNA\mathit{TFi}^n \cdot Ribosome]\end{aligned} \tag{7}$$

$$\begin{aligned}\frac{d[mRNA\mathit{TFi}^c \cdot Ribosome]}{dt} = &k_{A3,i}[mRNA\mathit{TFi}^c][Ribosome] - ... \\ &k_{D3,i}[mRNA\mathit{TFi}^c \cdot Ribosome] - ... \\ &k_{tl,i}[mRNA\mathit{TFi}^c \cdot Ribosome]\end{aligned} \tag{8}$$

$$\frac{d[Tfi^c]}{dt} = k_{tl,i}[mRNATFi^c \cdot Ribosome] - ... \\ -k^+_{nt,p}[Tfi^c] + k^-_{nt,p}[Tfi^n] - k_{d,p}[TFi^c] \quad (9)$$

$$\frac{d[Tfi^n]}{dt} = k^+_{nt,p}[Tfi^c] - k^-_{nt,p}[Tfi^n] - k_{d,p}[Tfi^n] - \\ \sum_{k=1}^{3}\Big(k_{A1,i,k}[TFk][Tfi^n] - k_{D1,i,k}[TFk \cdot Tfi^n]\Big) \quad (10)$$

Table 2. Kinetic parameters (approximate order of magnitude)

$constant$	deterministic	mesoscopic	reaction
$k^n_{A1,i,j}$	$10^9\ [M^{-1}\cdot min^{-1}]$	$1\ [\#^{-1}\cdot min^{-1}]$ *	association gene/protein
$k^n_{D1,i,j}$	$10^{-1}\ [min^{-1}]$	$10^{-1}\ [min^{-1}]$	dissociation gene/protein
$k^c_{A2,i,j}$	$10^8\ [M^{-1}\cdot min^{-1}]$	$10^{-2}\ [\#^{-1}\cdot min^{-1}]$ **	association protein/protein
$k^c_{D2,i,j}$	$10^{-1}\ [min^{-1}]$	$10^{-1}\ [min^{-1}]$	dissociation protein/protein
k_{tc}	$0.1\ [min^{-1}]$	$0.1\ [min^{-1}]$	transcription (catalytic step)
k_{tl}	$0.1\ [min^{-1}]$	$0.1\ [min^{-1}]$	translation (catalytic step)
$k_{d,p}$	$0.1\ [min^{-1}]$	$0.1\ [min^{-1}]$	decay rate proteins (catalytic step)
$k_{d,m}$	$0.1\ [min^{-1}]$	$0.1\ [min^{-1}]$	decay rate mRNA (catalytic step)
Assumed values for conversion: $V^n = 2fl$, * $V^c = 35fl$, ** $N_A = 6.022\cdot 10^{23}\#\cdot mol^{-1}$			

With initial conditions where the amount of the respective molecular species are actual integer numbers, a parametrization was performed to find a suitable steady state. The set of parameters thus obtained is certainly not unique, but it encompasses constants in a realistic order of magnitude. For the calculation of the deterministic time course, the MATLAB ode15s integrator was used to deal with possible stiffness of the system. The steady state shown in Fig. 2. corresponds to the exponential growth phase of the cell .

3 Mesoscopic Modelling

The deterministic approach considers average concentrations on a macroscopic level and is hence explicitly based on the assumption that there a is a sufficient number of molecules for each species. This does not hold in a single cell, with typically few copies of several species, in particular genes (see Sect. 2.). A continuous variation of the concentrations is physically not possible. The most accurate simulation would be on a microscopic level, considering molecular dynamics in terms of the actual positions and momenta of all molecules involved [GS02]. In between the two latter presentations lies the stochastic approach which deals with statistics and focuses only on the key molecules of the system. This level is called *mesoscopic* [VK83].

The *chemical master equation* is one possibility to solve chemical systems stochastically [MC67, EL01, VK83]. One here has to compute the probability distri-

bution that at an arbitrary time the species within a system are in a certain state. This basically means one has to find all feasible combinations of the molecules involved. When x_{max} is the largest molecule number in a system of *N* species, this means solving an *N*-dimensional cube with $\prod_{i=1}^{N}(x_{max}+1)$ grid points at each time step. One can increase the step size by applying a Fokker-Planck equation [ET88], however for a system in the present order of magnitude with 26 species, using a master equation will remain infeasible.

Kinetic Monte Carlo algorithms consider only a tractable number of species because they only deal with one overall state at each time step. Vice versa this means that one has to calculate several trajectories to explore the statistics of the system. Gillespie's Stochastic Simulation Algorithm (SSA) [GD77] was used here to solve the set of differential equations. Based on the reaction rate r_i the algorithm calculates the relative probability

$$p_i = \frac{r_i}{\sum_i r} \tag{11}$$

and then chooses, which step is next to occur. Additionally, the step length τ is determined via the equation

$$\tau = -\frac{log(p)}{\sum_i r}. \tag{12}$$

Hence, it is also stochastic. Note that one run of the SSA corresponds to a single point in the solution space of the chemical master equation and that for large number of molecules, this solution becomes equivalent to the result obtained with the deterministic approach. The drawback of the stochastic method is the possible stiffness of the model, analogous to the deterministic integration. This can be overcome by considering time steps that are sufficiently small, but it makes the problem computationally rather expensive.

4 Results

The original ODE system was decomposed to yield 48 elementary, irreversible reaction steps. The integer initial conditions of the deterministic integration were used for several SSA runs. The mean of these simulations was calculated and compared to the deterministic solution. Neglecting minor deviations, the deterministic result can be seen as the average of the SSA runs for the present model. The broad range of the stochastic solutions results in a large standard deviation for the respective components. (see Fig. 2.).

The system behavior is the same for both approaches. This was shown by testing the step response of the system upon a perturbation: At a certain point, the concentration of Tfm^c was lifted to a given value, to simulate for example a sudden signalling response in the cell. It can be seen, that the dynamic behavior of both simulations is also identical for this step response. Even more, the stochastic solution converges as shown for the mean of 25 and 250 SSA runs, respectively (see Fig. 3.).

The CPU time needed for a deterministic solution was about 1.1s while a single SSA run took about 4.5s on a *Pentium4* (2.4MHz) to simulate 300 min and reach a

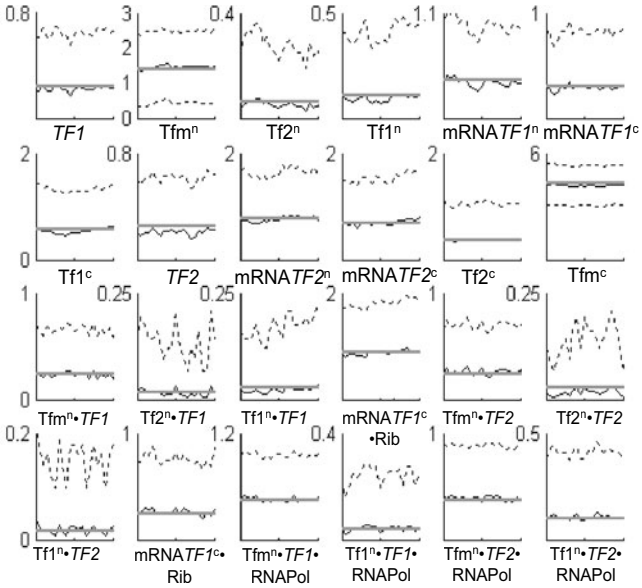


Fig. 2. Steady state for all system species over a time course of 10 minutes. The deterministic solution (solid grey lines) and the average of 250 SSA runs (solid black lines) as well as the respective standard deviations (dotted black lines) are shown.

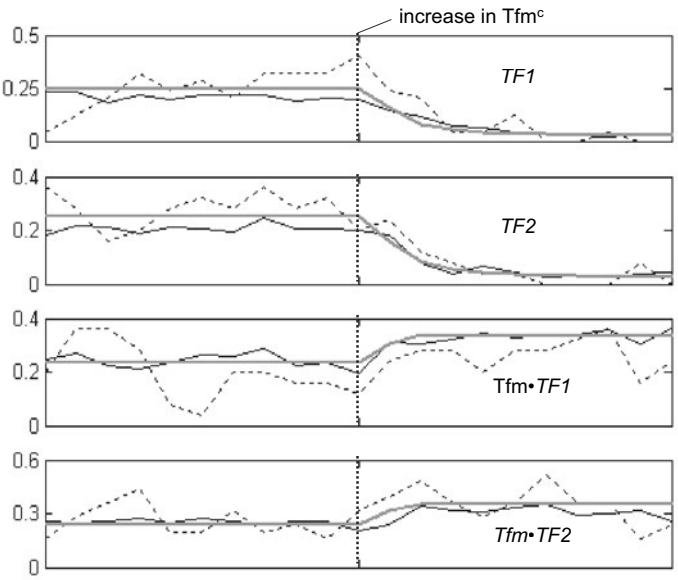


Fig. 3. Sample step response of 4 species upon a sudden increase of the tf_m concentration in the cytoplasm. The deterministic solution (solid grey lines) and the average of 25 (dotted black lines) and 250 SSA runs (solid black lines) are shown for a time course of 10 minutes.

steady state. Though this seems tolerable, one has to keep in mind, that the considered model represents a small example, which can be basically seen as a toy problem. When simulating bigger models of real regulatory networks consisting of (considerably) more than 10 genes, the pure SSA approach will quickly become infeasible. One possibility to speed up simulations might be the *tau leaping method*, which neglects the fast reactions for several time steps while focusing on the slow ones [RC03]. The general goal should be to keep the original ODE model completely unchanged to allow an automated application of stochastic differential equation (SDE) solvers on models written in *systems biology markup language* (SBML) [SB03].

Though the general agreement between deterministic and stochastic simulation is obvious in the present example problem (see Fig. 2.), one could think of cases, where the former approach would lead to a wrong outcome. If one considers for example a process in a noisy environment where the overall result is uncertain, SDEs become essential. A typical case would be whether a cell survives a virus infection or not, i.e. whether it undergoes lysis or lysogeny [MA98]. Another example are protein signalling cascades: The phenomenon of ultrasensitity switches is based on very small changes in the environment [FE99] and stochastics are very important here as well, especially if the overall system displays multiple steady states. Finally, gene expression itself is a noisy process [GG02, KE01, VL03]. Stochastics were shown to have for example a big impact on the heterogeneity of eucaryotic populations [BK03], a phenomenon that remains inexplicable if one considers pure deterministic simulations.

5 Summary/Conclusions

Expression of a two gene sample problem in an eucaryotic cell was modelled using elementary reactions. The original set of ODEs remained unchanged and no steady state assumptions were made. The model was simulated by means of deterministic and stochastic integration. SSA which was employed exhibited a reliable performance. Despite the obvious agreement between the deterministic and the stochastic approach, the latter presents a more accurate presentation of biological systems where the overall number of some molecular species is rather small. The drawback of stochastic simulation are the computational costs that are relatively high when compared to the deterministic integration. One possibility to overcome this restriction might be the *tau-leaping* method when considering larger systems for simulation and analysis.

Acknowledgements. The authors would like to thank Jörg Stelling of the Max-Planck-Institute of Complex Dynamical Systems in Magdeburg, Germany, for assistance and support on modelling issues.

6 Nomenclature

Tf, Tfi, Tfj, Tf1, Tf2, Tfm	proteins	
Tfa	activator protein	
TF, *TFi*, *TFj*, *TFk*, *TF1*, *TF2*	genes	
K_{eq}	equilibrium constant	[M, #]
N_A	Avogadro constant	$[\# \cdot min^{-1}]$
V	volume	[l]
k	kinetic rate constant	$[M^{-1} \cdot min^{-1}]$, $[\#^{-1} \cdot min^{-1}]$, $[min^{-1}]$
n	number of molecules	[#]
p	probability	[-]

Subscripts

A	association
D	dissociation
a	activator
d	decay
i,j,k	numbers of genes and proteins, respectively
m	mRNA
nt	nuclear transition
p	proteins
tc, tl	transription and translation, respectively

Superscripts

c	cytoplasm
n	nucleus
+	nuclear transition: cytoplasm → nucleus
-	nuclear transition: nucleus → cytoplasm

References

[BO86] Bailey, J.E., Ollis, D.F.: Biochemical Engineering Fundamentals. McGraw-Hill (1986)

[VK83] Kampen, van N.G.: Stochastic Processes in Physics and Chemistry. North-Holland Publishing Co., Amsterdam (1983)

[ET88] Erdi, P., Toth, J., Mathematical Models of chemical reactions. Princeton University Press, Princeton (1988)
[EL01] Elf, J., Lötstedt, P., Sjöberg, P.: Problems of high dimensions in molecular biology. In: Proceedings of the 17th GAMM-Seminar Leipzig 2001, 1–10 (2001)
[GD77] Gillespie, D.T., Exact Stochastic Simulation of Coupled Chemical Reactions. J. Phys. Chem., **81**, 2340ff (1977)
[MC67] McQuarrie, D.A., Stochastic Approach to Chemical Kinetics. J. Appl. Prob., **4**, 413-478 (1967)
[JB91] Bailey, J.E., Towards a science of metabolic engineering. Science, **252**, 1668-1674 (1991)
[RC03] Rathinam, M., Cao, Y., Petzold, L., Gillespie, D., Stiffness in Stochastic Chemically Reacting Systems: The Implicit Tau-Leaping Method, J. Chem. Phys., submitted,
[SBM03] Hucka et. al, The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinf. **19(4)**, 524-31 (2003)
[VK99] Varner J., Ramkrishna D., Mathematical models of metabolic pathways. Curr. Opin. Biotechnol. **10(2)**, 146-50 (1999)
[JB93] Bailey J.E., Host-vector interactions in Escherichia coli. Adv. Biochem. Eng. Biotechnol. **48**, 29-52 (1993)
[FE99] Ferrell, J.E.,Building a cellular switch: more lessons from a good egg. Bioessays. **21(10)**, 866-70 (1999)
[CO02] Cooper T.G., Transmitting the signal of excess nitrogen in Saccharomyces cerevisiae from the Tor proteins to the GATA factors: connecting the dots. FEMS Microbiol. Rev. **26(3)**, 223-38 (2002)
[GS02] Isralewitz B., Gao M., Schulten K., Steered molecular dynamics and mechanical functions of proteins. Curr. Opin. Struct. Biol. **11(2)**, 224-30 (2001)
[KE01] Kepler T.B., Elston T.C., Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys J. **81(6)**, 3116-36 (2001)
[GG02] Gonze D., Halloy J., Goldbeter A., Stochastic versus deterministic models for circadian rhythms. J. Biol. Phys. **28**, 637-53 (2002)
[SC03] Selinger D.W., Wright M.A., Church G.M., On the complete determination of biological systems. Trends Biotechnol. **21(6)**, 251-4 (2003)
[SH03] Stark J, Callard R, Hubank M., From the top down: towards a predictive biology of signalling networks. Trends Biotechnol. **21(7)**, 290-3 (2003)

Color Plates

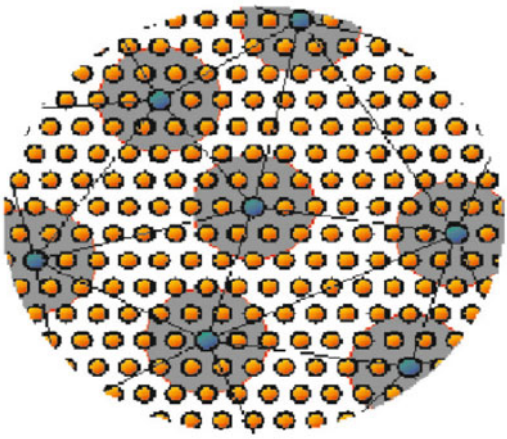


Plate 1. Schematic illustration of QC (courtesy of M. Ortiz). Only atoms in the small cluster need to be visited during the computation. (See Fig. 1 on Page 6)

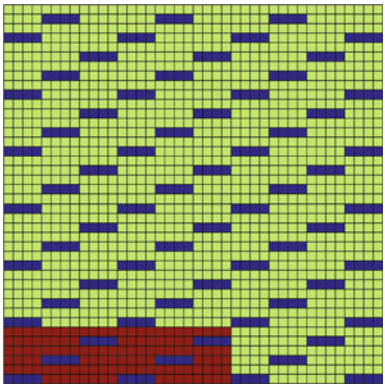


Plate 2. Sample of the microarray. (See Fig. 1 on Page 29)

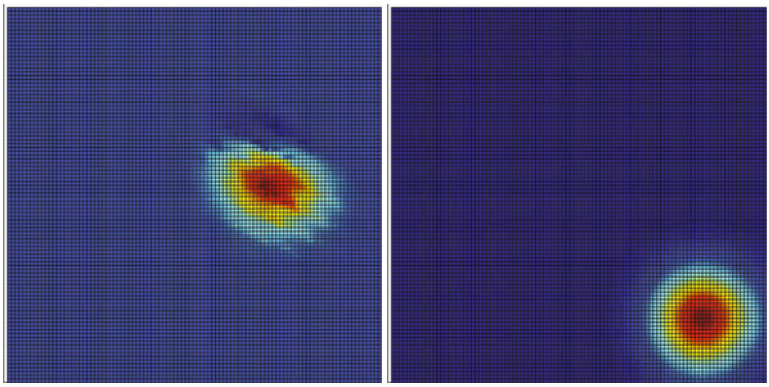


Plate 3. Transport simulation in a homogeneous array versus heterogeneous array, $k = 0.1$, $D = 0.0005$. (See Fig. 2 on Page 31)

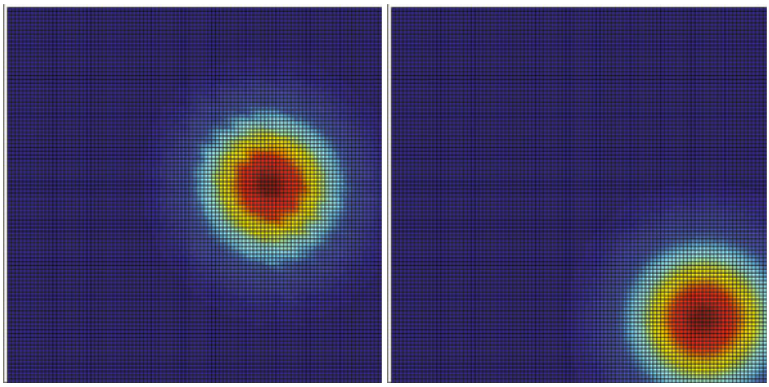


Plate 4. Transport simulation in a homogeneous array versus heterogeneous array, $k = 0.1$, $D = 0.001$. (See Fig. 3 on Page 31)

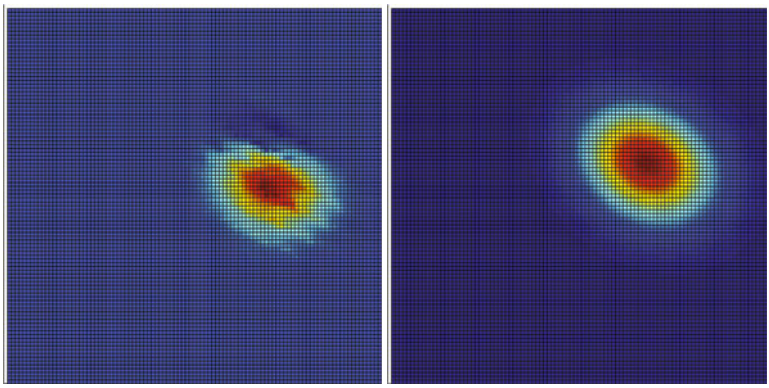


Plate 5. Transport simulation for the homogenized array versus heterogeneous array, $D = 0.0005$. (See Fig. 4 on Page 32)

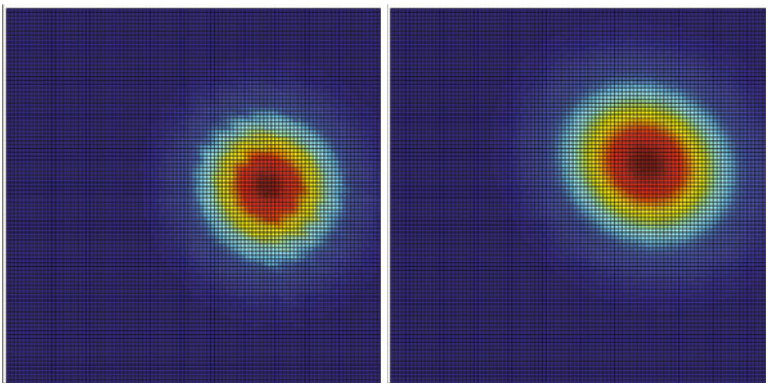


Plate 6. Transport simulation for the homogenized array versus heterogeneous array, $D = 0.001$. (See Fig. 5 on Page 32)

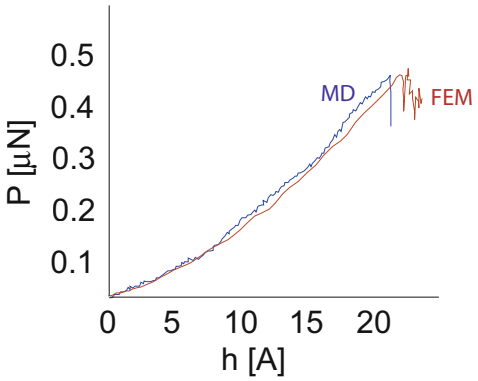


Plate 7. Nanoindentation displacement-load (h-P) response showing initial elastic deformation of the thin film up to a critical indentation distance where the load suddenly drops (signaling the nucleation of a dislocation loop), prediction by FEM with Cauchy-Born hypothesis (smooth curve) and direct MD simulation (wiggly curve). (See Fig. 2 on Page 73)

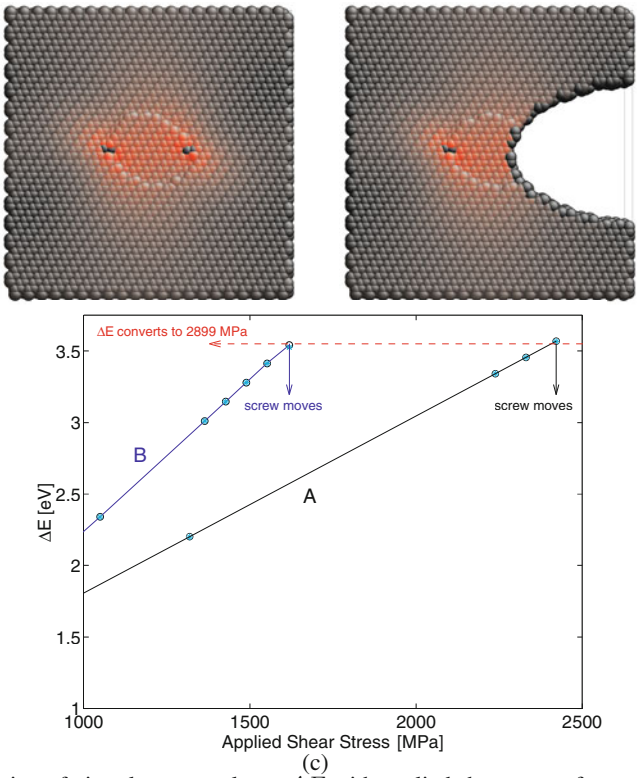


Plate 8. Variation of virtual energy release ΔE with applied shear stress for a screw dislocation in silicon in two different microstructure environments (A and B), one in the form of a dislocation dipole, two dislocation cores as shown in (a), and the other in the form of a dislocation core with a notch nearby, as shown in (b). The left dislocation core is seen to to move at the same value of ΔE but different values of the applied shear stress. This signifies ΔE is not affected by the samepl microstructure environment and therefore it can be converted to a value for the Peierls stress that is an intrinsic material property. (See Fig. 3 on Page 75)

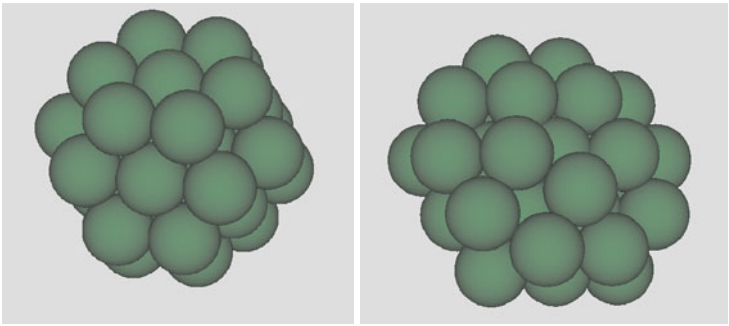


Plate 9. The two lowest energy minima for the 38-atom Lennard-Jones cluster: a face centered cubic minimum (left) and a Mackay icosahedron (right). (See Fig. 1 on Page 83)

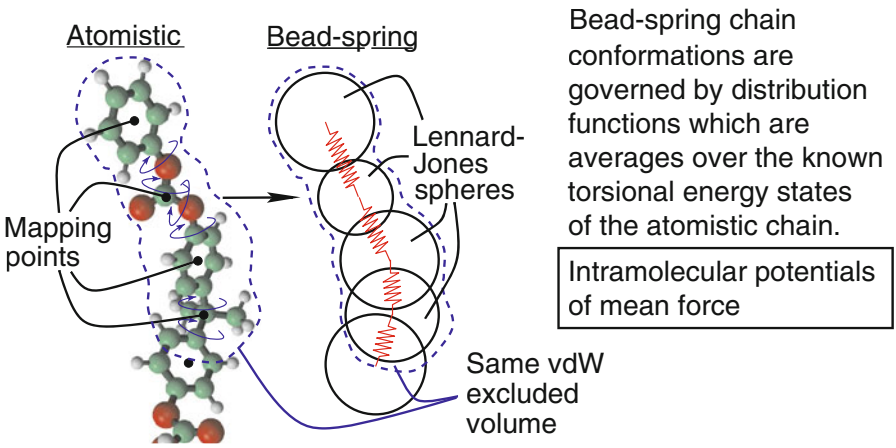


Plate 10. A schematic representation of the 4:1 mapping scheme for coarse-graining of bisphenol-A-polycarbonate (BPA-PC). A chain-ending repeat unit is shown. From [Ab03]. (See Fig. 5 on Page 114)

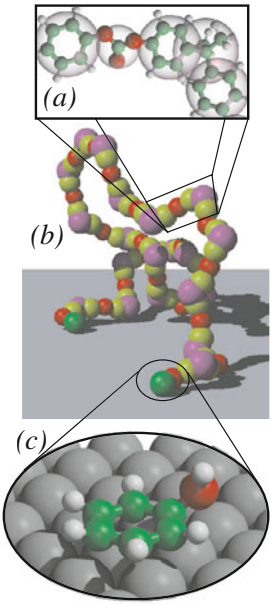


Plate 11. Pictorial illustration of the multiscale model for BPA-PC: (a) coarse-grained beads (transparent spheres) superimposed on the underlying chemical structure; (b) Coarse-grained model of an $N = 20$ BPA-PC molecule, with ends adsorbed on the surface. (c) A phenol molecule adsorbed on the bridge site of a Ni (111) surface; configuration computed via CPMD simulation. (From Ref.[4]) (See Fig. 8 on Page 127)

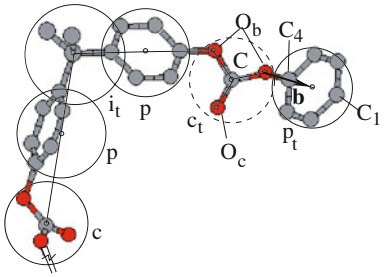


Plate 12. Atomistic structure of a representative chain-end conformation of BPA-PC illustrating the dual-resolution scheme. The small black circles are mapping points in the scheme, and the large circles represent the excluded volume diameters of those points. The C_4-C_1 orientation vector, $\mathbf{b}$, is shown. Lower-case letters denote coarse-grained bead designations: "p" is phenylene, "c_t" is the terminal carbonate, "p_t" is the terminal phenoxy, and "i_t" is the terminal isopropylidene. Capital letters denote atom type designations (see text). (See Fig. 2 on Page 136)

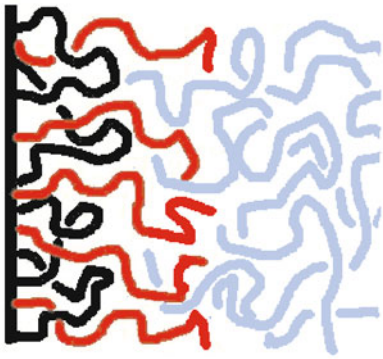


Plate 13. Schematic representation of the molecular structure of a BPA-PC melt adjacent to a Ni wall, showing double-ended adsorbed, single-ended adsorbed, and non-adsorbed chains. (See Fig. 4 on Page 139)

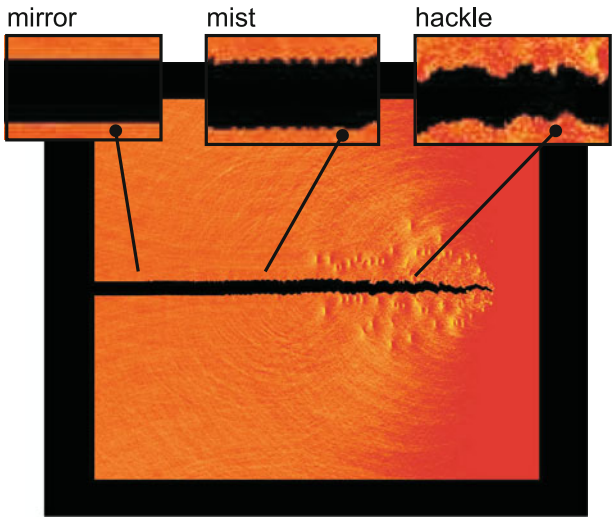


Plate 14. Crack propagation in a LJ system as reported earlier in [ABRR94]. The plot shows the σ_{xx}-field and indicates the mirror-mist-hackle transition. (See Fig. 1 on Page 145)

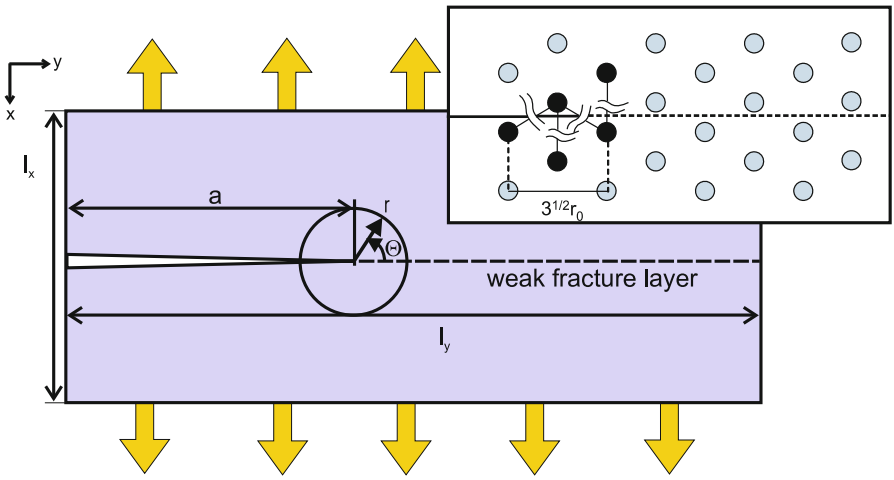


Plate 15. Geometry of the simulation slab. The slab width is denoted by l_x, the slab length is l_y and the crack extension is a. The plot also depicts the crack orientation and the bond breaking process at the crack tip. (See Fig. 2 on Page 146)

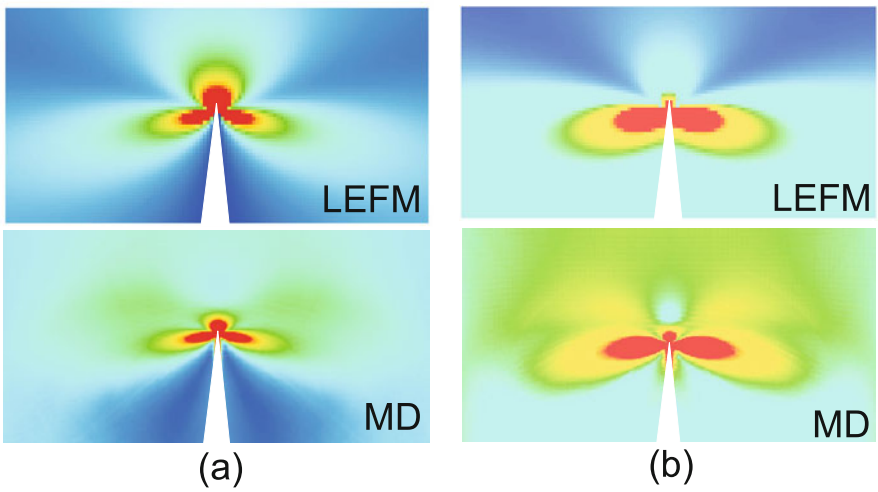


Plate 16. Potential energy field and magnitude of the dynamic Poynting vector. (a) Potential energy field near a crack close to the Rayleigh speed. (b) Energy flow near a rapidly propagating crack. This plot shows the magnitude of the dynamic Poynting vector in the vicinity of a crack propagating at a velocity close to the Rayleigh speed. (See Fig. 9 on Page 152)

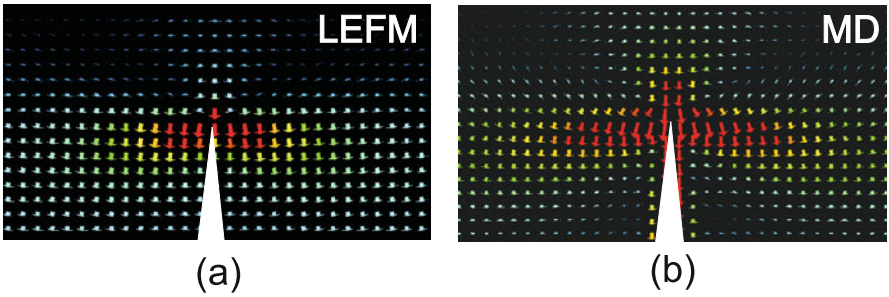


Plate 17. Energy flow near a rapidly propagating crack. This plot shows (a) the continuum mechanics prediction, and (b) the MD simulation result of the dynamic Poynting vector field in the vicinity of the crack tip, for a crack propagating close to the Rayleigh speed. (See Fig. 10 on Page 153)

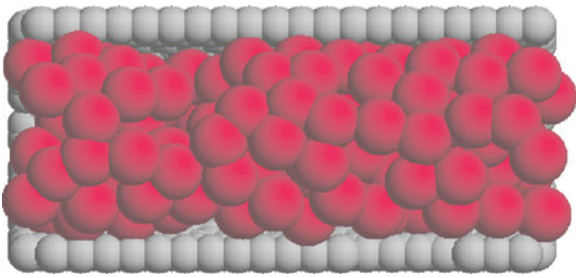


Plate 18. Snapshot from NEMD simulations of the flow of methane inside a (16,16) carbon nanotube. Half of the carbon nanotube has been removed to view the methane molecules. (See Fig. 2 on Page 221)

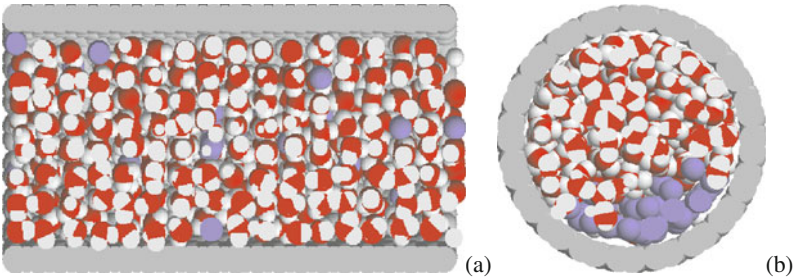


Plate 19. Snapshots of the initial molecular configuration (a) and the equilibrated system (b) of a water-nitrogen mixture flowing inside a carbon nanotube. (See Fig. 6 on Page 223)

Editorial Policy

§1. Volumes in the following three categories will be published in LNCSE:

i) Research monographs
ii) Lecture and seminar notes
iii) Conference proceedings

Those considering a book which might be suitable for the series are strongly advised to contact the publisher or the series editors at an early stage.

§2. Categories i) and ii). These categories will be emphasized by Lecture Notes in Computational Science and Engineering. **Submissions by interdisciplinary teams of authors are encouraged.** The goal is to report new developments – quickly, informally, and in a way that will make them accessible to non-specialists. In the evaluation of submissions timeliness of the work is an important criterion. Texts should be well-rounded, well-written and reasonably self-contained. In most cases the work will contain results of others as well as those of the author(s). In each case the author(s) should provide sufficient motivation, examples, and applications. In this respect, Ph.D. theses will usually be deemed unsuitable for the Lecture Notes series. Proposals for volumes in these categories should be submitted either to one of the series editors or to Springer-Verlag, Heidelberg, and will be refereed. A provisional judgment on the acceptability of a project can be based on partial information about the work: a detailed outline describing the contents of each chapter, the estimated length, a bibliography, and one or two sample chapters – or a first draft. A final decision whether to accept will rest on an evaluation of the completed work which should include

– at least 100 pages of text;
– a table of contents;
– an informative introduction perhaps with some historical remarks which should be accessible to readers unfamiliar with the topic treated;
– a subject index.

§3. Category iii). Conference proceedings will be considered for publication provided that they are both of exceptional interest and devoted to a single topic. One (or more) expert participants will act as the scientific editor(s) of the volume. They select the papers which are suitable for inclusion and have them individually refereed as for a journal. Papers not closely related to the central topic are to be excluded. Organizers should contact Lecture Notes in Computational Science and Engineering at the planning stage.

In exceptional cases some other multi-author-volumes may be considered in this category.

§4. Format. Only works in English are considered. They should be submitted in camera-ready form according to Springer-Verlag's specifications.
Electronic material can be included if appropriate. Please contact the publisher.
Technical instructions and/or TEX macros are available via
http://www.springeronline.com/sgw/cda/frontpage/0,10735,5-111-2-71391-0,00.html
The macros can also be sent on request.

General Remarks

Lecture Notes are printed by photo-offset from the master-copy delivered in camera-ready form by the authors. For this purpose Springer-Verlag provides technical instructions for the preparation of manuscripts. See also *Editorial Policy*.

Careful preparation of manuscripts will help keep production time short and ensure a satisfactory appearance of the finished book.

The following terms and conditions hold:

Categories i), ii), and iii):
Authors receive 50 free copies of their book. No royalty is paid. Commitment to publish is made by letter of intent rather than by signing a formal contract. Springer-Verlag secures the copyright for each volume.

For conference proceedings, editors receive a total of 50 free copies of their volume for distribution to the contributing authors.

All categories:
Authors are entitled to purchase further copies of their book and other Springer mathematics books for their personal use, at a discount of 33,3 % directly from Springer-Verlag.

Addresses:

Timothy J. Barth
NASA Ames Research Center
NAS Division
Moffett Field, CA 94035, USA
e-mail: barth@nas.nasa.gov

Michael Griebel
Institut für Angewandte Mathematik
der Universität Bonn
Wegelerstr. 6
53115 Bonn, Germany
e-mail: griebel@iam.uni-bonn.de

David E. Keyes
Department of Applied Physics
and Applied Mathematics
Columbia University
200 S. W. Mudd Building
500 W. 120th Street
New York, NY 10027, USA
e-mail: david.keyes@columbia.edu

Risto M. Nieminen
Laboratory of Physics
Helsinki University of Technology
02150 Espoo, Finland
e-mail: rni@fyslab.hut.fi

Dirk Roose
Department of Computer Science
Katholieke Universiteit Leuven
Celestijnenlaan 200A
3001 Leuven-Heverlee, Belgium
e-mail: dirk.roose@cs.kuleuven.ac.be

Tamar Schlick
Department of Chemistry
Courant Institute of Mathematical
Sciences
New York University
and Howard Hughes Medical Institute
251 Mercer Street
New York, NY 10012, USA
e-mail: schlick@nyu.edu

Springer-Verlag, Mathematics Editorial IV
Tiergartenstrasse 17
69121 Heidelberg, Germany
Tel.: *49 (6221) 487-8185
e-mail: peters@springer.de

Lecture Notes in Computational Science and Engineering

Vol. 1 D. Funaro, *Spectral Elements for Transport-Dominated Equations.* 1997. X, 211 pp. Softcover. ISBN 3-540-62649-2

Vol. 2 H. P. Langtangen, *Computational Partial Differential Equations.* Numerical Methods and Diffpack Programming. 1999. XXIII, 682 pp. Hardcover. ISBN 3-540-65274-4

Vol. 3 W. Hackbusch, G. Wittum (eds.), *Multigrid Methods V.* Proceedings of the Fifth European Multigrid Conference held in Stuttgart, Germany, October 1-4, 1996. 1998. VIII, 334 pp. Softcover. ISBN 3-540-63133-X

Vol. 4 P. Deuflhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, R. D. Skeel (eds.), *Computational Molecular Dynamics: Challenges, Methods, Ideas.* Proceedings of the 2nd International Symposium on Algorithms for Macromolecular Modelling, Berlin, May 21-24, 1997. 1998. XI, 489 pp. Softcover. ISBN 3-540-63242-5

Vol. 5 D. Kröner, M. Ohlberger, C. Rohde (eds.), *An Introduction to Recent Developments in Theory and Numerics for Conservation Laws.* Proceedings of the International School on Theory and Numerics for Conservation Laws, Freiburg / Littenweiler, October 20-24, 1997. 1998. VII, 285 pp. Softcover. ISBN 3-540-65081-4

Vol. 6 S. Turek, *Efficient Solvers for Incompressible Flow Problems.* An Algorithmic and Computational Approach. 1999. XVII, 352 pp, with CD-ROM. Hardcover. ISBN 3-540-65433-X

Vol. 7 R. von Schwerin, ***M**ulti **B**ody **S**ystem **SIM**ulation.* Numerical Methods, Algorithms, and Software. 1999. XX, 338 pp. Softcover. ISBN 3-540-65662-6

Vol. 8 H.-J. Bungartz, F. Durst, C. Zenger (eds.), *High Performance Scientific and Engineering Computing.* Proceedings of the International FORTWIHR Conference on HPSEC, Munich, March 16-18, 1998. 1999. X, 471 pp. Softcover. 3-540-65730-4

Vol. 9 T. J. Barth, H. Deconinck (eds.), *High-Order Methods for Computational Physics.* 1999. VII, 582 pp. Hardcover. 3-540-65893-9

Vol. 10 H. P. Langtangen, A. M. Bruaset, E. Quak (eds.), *Advances in Software Tools for Scientific Computing.* 2000. X, 357 pp. Softcover. 3-540-66557-9

Vol. 11 B. Cockburn, G. E. Karniadakis, C.-W. Shu (eds.), *Discontinuous Galerkin Methods.* Theory, Computation and Applications. 2000. XI, 470 pp. Hardcover. 3-540-66787-3

Vol. 12 U. van Rienen, *Numerical Methods in Computational Electrodynamics.* Linear Systems in Practical Applications. 2000. XIII, 375 pp. Softcover. 3-540-67629-5

Vol. 13 B. Engquist, L. Johnsson, M. Hammill, F. Short (eds.), *Simulation and Visualization on the Grid.* Parallelldatorcentrum Seventh Annual Conference, Stockholm, December 1999, Proceedings. 2000. XIII, 301 pp. Softcover. 3-540-67264-8

Vol. 14 E. Dick, K. Riemslagh, J. Vierendeels (eds.), *Multigrid Methods VI.* Proceedings of the Sixth European Multigrid Conference Held in Gent, Belgium, September 27-30, 1999. 2000. IX, 293 pp. Softcover. 3-540-67157-9

Vol. 15 A. Frommer, T. Lippert, B. Medeke, K. Schilling (eds.), *Numerical Challenges in Lattice Quantum Chromodynamics.* Joint Interdisciplinary Workshop of John von Neumann Institute for Computing, Jülich and Institute of Applied Computer Science, Wuppertal University, August 1999. 2000. VIII, 184 pp. Softcover. 3-540-67732-1

Vol. 16 J. Lang, *Adaptive Multilevel Solution of Nonlinear Parabolic PDE Systems.* Theory, Algorithm, and Applications. 2001. XII, 157 pp. Softcover. 3-540-67900-6

Vol. 17 B. I. Wohlmuth, *Discretization Methods and Iterative Solvers Based on Domain Decomposition.* 2001. X, 197 pp. Softcover. 3-540-41083-X

Vol. 18 U. van Rienen, M. Günther, D. Hecht (eds.), *Scientific Computing in Electrical Engineering.* Proceedings of the 3rd International Workshop, August 20-23, 2000, Warnemünde, Germany. 2001. XII, 428 pp. Softcover. 3-540-42173-4

Vol. 19 I. Babuška, P. G. Ciarlet, T. Miyoshi (eds.), *Mathematical Modeling and Numerical Simulation in Continuum Mechanics.* Proceedings of the International Symposium on Mathematical Modeling and Numerical Simulation in Continuum Mechanics, September 29 - October 3, 2000, Yamaguchi, Japan. 2002. VIII, 301 pp. Softcover. 3-540-42399-0

Vol. 20 T. J. Barth, T. Chan, R. Haimes (eds.), *Multiscale and Multiresolution Methods.* Theory and Applications. 2002. X, 389 pp. Softcover. 3-540-42420-2

Vol. 21 M. Breuer, F. Durst, C. Zenger (eds.), *High Performance Scientific and Engineering Computing.* Proceedings of the 3rd International FORTWIHR Conference on HPSEC, Erlangen, March 12-14, 2001. 2002. XIII, 408 pp. Softcover. 3-540-42946-8

Vol. 22 K. Urban, *Wavelets in Numerical Simulation.* Problem Adapted Construction and Applications. 2002. XV, 181 pp. Softcover. 3-540-43055-5

Vol. 23 L. F. Pavarino, A. Toselli (eds.), *Recent Developments in Domain Decomposition Methods.* 2002. XII, 243 pp. Softcover. 3-540-43413-5

Vol. 24 T. Schlick, H. H. Gan (eds.), *Computational Methods for Macromolecules: Challenges and Applications.* Proceedings of the 3rd International Workshop on Algorithms for Macromolecular Modeling, New York, October 12-14, 2000. 2002. IX, 504 pp. Softcover. 3-540-43756-8

Vol. 25 T. J. Barth, H. Deconinck (eds.), *Error Estimation and Adaptive Discretization Methods in Computational Fluid Dynamics.* 2003. VII, 344 pp. Hardcover. 3-540-43758-4

Vol. 26 M. Griebel, M. A. Schweitzer (eds.), *Meshfree Methods for Partial Differential Equations*. 2003. IX, 466 pp. Softcover. 3-540-43891-2

Vol. 27 S. Müller, *Adaptive Multiscale Schemes for Conservation Laws*. 2003. XIV, 181 pp. Softcover. 3-540-44325-8

Vol. 28 C. Carstensen, S. Funken, W. Hackbusch, R. H. W. Hoppe, P. Monk (eds.), *Computational Electromagnetics*. Proceedings of the GAMM Workshop on "Computational Electromagnetics", Kiel, Germany, January 26-28, 2001. 2003. X, 209 pp. Softcover. 3-540-44392-4

Vol. 29 M. A. Schweitzer, *A Parallel Multilevel Partition of Unity Method for Elliptic Partial Differential Equations*. 2003. V, 194 pp. Softcover. 3-540-00351-7

Vol. 30 T. Biegler, O. Ghattas, M. Heinkenschloss, B. van Bloemen Waanders (eds.), *Large-Scale PDE-Constrained Optimization*. 2003. VI, 349 pp. Softcover. 3-540-05045-0

Vol. 31 M. Ainsworth, P. Davies, D. Duncan, P. Martin, B. Rynne (eds.) *Topics in Computational Wave Propagation*. Direct and Inverse Problems. 2003. VIII, 399 pp. Softcover. 3-540-00744-X

Vol. 32 H. Emmerich, B. Nestler, M. Schreckenberg (eds.) *Interface and Transport Dynamics*. Computational Modelling. 2003. XV, 432 pp. Hardcover. 3-540-40367-1

Vol. 33 H. P. Langtangen, A. Tveito (eds.) *Advanced Topics in Computational Partial Differential Equations*. Numerical Methods and Diffpack Programming. 2003. XIX, 658 pp. Softcover. 3-540-01438-1

Vol. 34 V. John, *Large Eddy Simulation of Turbulent Incompressible Flows*. Analytical and Numerical Results for a Class of LES Models. 2004. XII, 261 pp. Softcover. 3-540-40643-3

Vol. 35 E. Bänsch, *Challenges in Scientific Computing - CISC 2002*. Proceedings of the Conference *Challenges in Scientific Computing*, Berlin, October 2-5, 2002. 2003. VIII, 287 pp. Hardcover. 3-540-40887-8

Vol. 36 B. N. Khoromskij, G. Wittum, *Numerical Solution of Elliptic Differential Equations by Reduction to the Interface*. 2004. XI, 293 pp. Softcover. 3-540-20406-7

Vol. 37 A. Iske, *Multiresolution Methods in Scattered Data Modelling*. 2004. XII, 182 pp. Softcover. 3-540-20479-2

Vol. 38 S.-I. Niculescu, K. Gu, *Advances in Time-Delay Systems*. 2004. XIV, 446 pp. Softcover. 3-540-20890-9

Vol. 39 S. Attinger, P. Koumoutsakos, *Multiscale Modelling and Simulation*. 2004. VIII, 277 pp. Softcover. 3-540-21180-2

For further information on these books please have a look at our mathematics catalogue at the following URL: `www.springeronline.com/series/3527`

Texts in Computational Science and Engineering

Vol. 1 H. P. Langtangen, *Computational Partial Differential Equations.* Numerical Methods and Diffpack Programming. 2nd Edition 2003. XXVI, 855 pp. Hardcover. ISBN 3-540-43416-X

Vol. 2 A. Quarteroni, F. Saleri, *Scientific Computing with MATLAB.* 2003. IX, 257 pp. Hardcover. ISBN 3-540-44363-0

Vol. 3 H. P. Langtangen, *Python Scripting for Computational Science.* 2004. XXII, 724 pp. Hardcover. ISBN 3-540-43508-5

For further information on these books please have a look at our mathematics catalogue at the following URL: `www.springeronline.com/series/5151`

Printing and Binding: Strauss GmbH, Mörlenbach