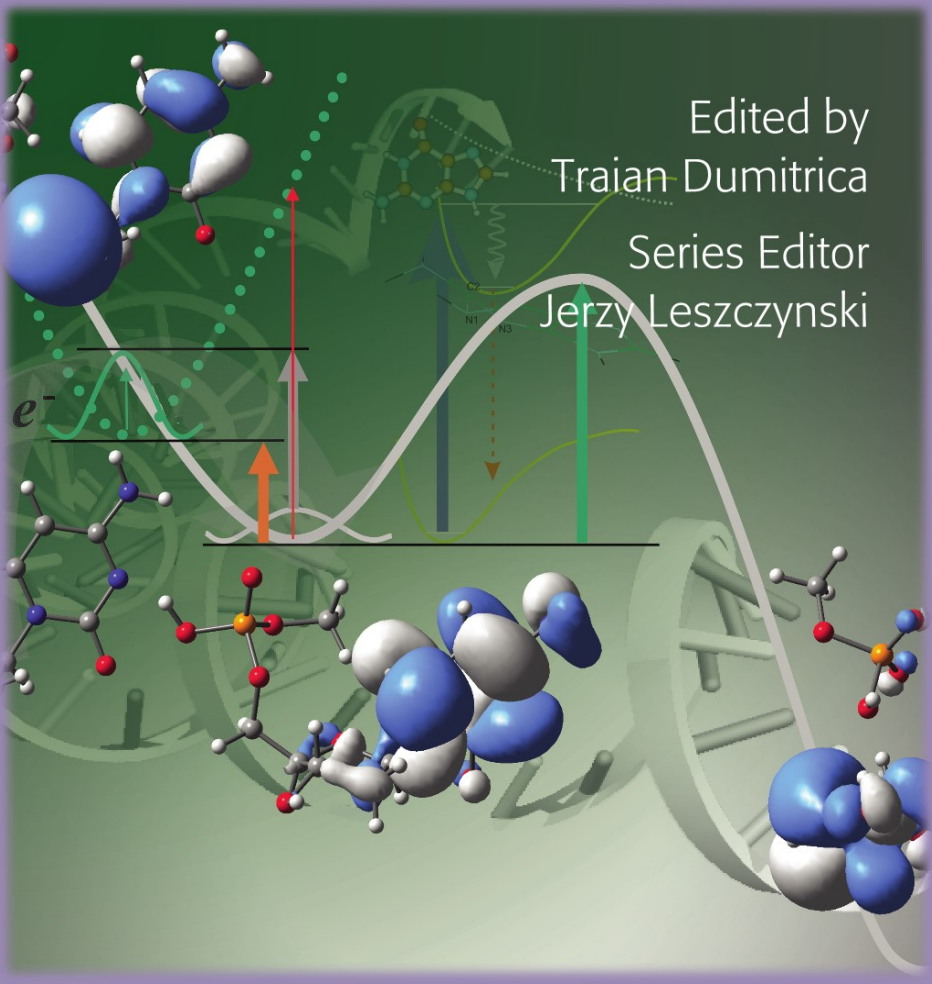
CHALLENGES AND ADVANCES IN COMPUTATIONAL CHEMISTRY AND PHYSICS

9

Trends in Computational Nanomechanics

Transcending Length and Time Scales



Springer

Trends in Computational Nanomechanics

CHALLENGES AND ADVANCES IN COMPUTATIONAL CHEMISTRY AND PHYSICS

Volume 9

Series Editor:

JERZY LESZCZYNSKI
Department of Chemistry, Jackson State University, U.S.A.

For further volumes:
http://www.springer.com/series/6918

Trends in Computational Nanomechanics

Transcending Length and Time Scales

Edited by

Traian Dumitrica
University of Minnesota, Minneapolis, USA

Springer

Editor
Dr. Traian Dumitrica
University of Minnesota
Department of Mechanical Engineering
111 Church St. SE.,
Minneapolis MN 55455
USA
td@me.umn.edu

ISBN 978-1-4020-9784-3 e-ISBN 978-1-4020-9785-0
DOI 10.1007/978-1-4020-9785-0
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009939460

PREFACE

Tubes, rods, plates, shells, etc., are ubiquitous engineering structures, traditionally imagined as continuous objects. Their mechanical behaviour is usually studied with the computational engineering tools of the continuum scale, based on the relations of macroscopic elasticity. These methods are computationally efficient, since they do not need to track every single atom. Thermodynamic quantities, such as temperature, are represented as fields. Behind this static elastic continuum is the effervescent world of mechanics – *Nanomechanics* – where temperature represents the kinetic energy of the random motion of an order of 10^{23} microscopic entities. This dynamical discrete behaviour shapes the experimentally verifiable constitutive laws of the continuum.

Nano-tubes, -wires, -coils, and -plates are organizations of matter that became possible with recent advances in synthetic methods. In these novel structures, macroscopic mechanics breaks down and nanomechanical behaviour emerges. Understanding the new mechanical response is important both fundamentally and practically. By capitalizing on the science emerging from the newly accessible size range, engineers can develop electromechanical devices, machines, and electronics on the nano scale. In spite of a vast body of research, however, the inherent practical difficulties at the nano scale have prevented most applications from being achieved. This is a powerful motivation for computational modeling, which is increasingly viewed as a third technique, complementing theory and experiment.

Situated at the intersection of computational chemistry, solid state physics, and mechanical engineering, *computational nanomechanics* has emerged as an exciting area of research that has already played an important role in understanding complex mechanical responses at the nano scale. Important nanomechanical problems can be simulated with traditional computational approaches, such as molecular dynamics (for the microscopic scale) and finite elements (for the continuum scale). Many other important problems (e.g., the overall strength of a material) span large time and size scales and are strongly influenced by atomic-scale processes (e.g., fracture propagation). Performing an exhaustive simulation at the smallest/fastest scale for a domain of engineering significance is computationally prohibitive and likely to remain so for a very long time. Recent advances in computational methodologies have made it possible to go beyond the distinct approaches mentioned above. By linking previously "separated by discipline" investigation tools, multi-scale nanomechanical aspects begin to be simulated and studied from both fundamental and engineering-application viewpoints.

Trends in Computational Nanomechanics is an informative description of the progress in nanomechanics as of 2009. It captures the important insights generated via the separate application of microscopic (from ab initio to tight binding to empirical force field) and continuum modeling techniques, as well as recent developments in multiscale methodologies and the new results generated by utilizing these methods. The book is a useful reference for graduates, undergraduates, and professionals interested in this interdisciplinary research area.

Nanomechanical problems involving large-scale phenomena are particularly suitable for *hybrid* multiscale numerical modeling strategies. Hybrid schemes combine within the same overall spatial and temporal domain, several physical models that are separately designed to address different length and time scales. The assortment of models can have similar or fundamentally different ways of representing the physical processes. For example, several schemes couple molecular dynamics simulations where the atomistic system is described with quantum mechanical and classical potential models, whereas others couple classical molecular dynamics at the microscopic scale to continuum mechanics (typically finite element) simulations at larger scales. The main difficulty with developing hybrid schemes is reconciling the boundaries between the domains represented with the various models.

When various regions of a nanomechanical object are treated with diverse atomistic resolutions, differences arise because of the inherent differences in the various levels of description. In this respect, Chapter 1, contributed by A. de Vita et al., presents the "Learn on The Fly" hybrid quantum/classical molecular dynamics. The merit of this fully atomistic scheme is that it successfully washes out the differences between the two atomistic-level models by incorporating adjustable classical potentials. Using it, one can investigate the motion of dislocations, brittle fracture, and growth of hydrogen-induced platelets in silicon. Hybrid molecular dynamics is also the focus of Chapter 2. Encouraged by the success of the hybrid quantum/classical molecular dynamics scheme, B. Ensing and S.O. Nielsen take an important step toward reaching the mesoscale by developing atomistic/coarse-grained molecular dynamics. Each coarse grain in their method lumps on the order of ten atoms. This chapter addresses in depth the technical details of the coupling between particles in the different resolution regions and the treatment of particles that cross a boundary between regions of different resolution and offers case studies.

Conventional molecular dynamics places severe limitations on the time scale afforded by the simulation. Recently, significant activities have focused on time-accelerated molecular dynamics methods that aim to push the limits of the simulation time and thus eliminate the gulf between nanomechanical simulation and experiment. Chapter 3, contributed by M. Grünwald and C. Dellago, discusses transition path sampling, a set of computational methods designed to overcome the timescale problem. The relevance of this method for nanomechanics is well illustrated with a study of the structural transformations of nanocrystals under pressure.

Microscopic methods are attractive starting points for understanding nanoscale thermal transport. Chapter 4, contributed by S.P.A. Gill, reviews molecular dynamics modeling methodologies suitable for conducting concurrent multiscale

simulations of heat conduction in solids, the fundamental connections between the thermophysical properties of a material and the quantum model of phonons, and the progress in atomistic modeling in heat transport in relation to nonequilibrium molecular dynamics. On the same theme, Chapter 5, by I.K. Puri and S. Murad, presents a hybrid methodology to couple molecular dynamics with the mesoscale lattice Boltzmann method.

Due to the increased importance of surface effects, the behaviour and properties of nano-objects generally differ from those of their bulk counterparts. To understand these new behaviours, combining hierarchically atomistic treatments with continuum methods is an attractive strategy. In this respect, Chapter 6, contributed by V. Dupont and F. Sansoz, presents the insight gained by applying large scale parallel molecular dynamics and the quasicontinuum method in contact-induced plasticity studies in nanocrystalline metals. Chapter 7, contributed by R.W. Nunes and J.F. Justo, provides an overview of the microscopic studies of silicon nanowires. To account for the larger scale, the authors propose extracting scaling laws for the energetics of nanowire motifs exposing surfaces. Chapter 8, contributed by H.S. Park and P.A. Klein, discusses the methodology to incorporate surface effects in a continuum framework and thus to efficiently model nanowires in engineering applications. Chapter 9, by C.V. Ciobanu et al., discusses the use of genetic computational methodologies to obtain optimal surface reconstruction and morphologies in nanowires.

Carbon nanotubes continue to be the focus of intense investigations. Significant insights into the behavior of these long molecules, or small solids, have been obtained with classical potential atomistic treatments. Chapter 10, by B.-W. Jeong and S.B. Sinnott, discusses molecular dynamics simulation methods and their applicability in studying the resilient response of carbon nanotube systems to compressive and torsional loads. A detailed treatment of covalent bonding limits the scale range covered by molecular dynamics. Chapter 11, by T. Chang, demonstrates that specific aspects of the mechanical response of carbon nanotubes at different length scales can be obtained analytically, starting from very simplified forms of the interatomic potentials, such as harmonic and Morse-type. Due to their high axial stiffness, nanotubes possess a macroscopic persistent length. As a result, nanotubes can be efficiently glued together over the extended contacts through van der Waals interactions, to form ropes, rings, and cables. In this respect, Chapter 12, contributed by Y. Huang et al., describes how to efficiently account for the van der Waals interlayer interactions of graphitic structures in molecular dynamics simulations.

Hierarchical and hybrid multiscale models are extremely useful in studying the rich mechanics of large-size carbon nanotubes. Chapter 13, by A. Pantano, presents a computationally-effective hierarchical tight-binding finite element scheme that can simulate the electromechanical behaviour of nanotubes of realistic dimensions. Chapter 14, by Y. Sun and K.M. Liew, provides insight into several atomistic to continuum coupling schemes and focuses on the coupling of a particular atomistic method to a higher-order continuum model. Chapter 15, by S. Im et al., presents a method that embeds quantum mechanics into a molecular mechanics quasicontinuum in order to simulate the mechanical deformation of curved crystalline bodies.

The effectiveness of this hybrid method is demonstrated in several situations that describe very well the mechanics of carbon nanotubes. Chapter 16, by Y.A. Wang et al., surveys commonly employed first-principles theoretical methods and focuses on the popular hybrid ONIOM scheme to determine the electronic properties and reactivity of perfect, defective, and doped carbon nanotubes.

The multiscale methodologies developed for nanomechanics show a remarkable transferability to the more complex area of biomaterials. Their application in studying the biomechanical response, which is the focus of the last three chapters of the volume, is currently a frontier research area called Meteriomics. Chapter 17, by M.J. Buehler et al., reviews the relevant atomistic modeling approaches and discusses the size-dependent fracture behaviour of protein materials via the application of hybrid multiscale strategies linking atomistic and continuum scales. Chapter 18, by X. Chen and Q. Cui, illustrates the utility of hierarchical modeling in understanding the fundamental processes of large channel conductance in Escherichia coli bacteria. The networks of biopolymers within a tissue exhibit complex multiscale mechanics, governed by the properties of the individual entities and the relationship between them. Chapter 19, by V.H. Barocas et al., reviews the current multiscale models developed for tissues and critically examines their validity against experimental results. This chapter demonstrates that multiscale modeling of networks has the potential to address some of the fundamental questions in the mechanobiology of tissues.

I thank the authors for the quality and completeness of their chapters, which highlight both the advances and many of the remaining challenges in this area. I also thank the reviewers for their stimulating comments and Ilia Nikiforov for help with the manuscript.

Minneapolis, MN Traian Dumitrica
June 2009

CONTENTS

CHAPTER 1

HYBRID QUANTUM/CLASSICAL MODELING OF MATERIAL SYSTEMS: THE "LEARN ON THE FLY" MOLECULAR DYNAMICS SCHEME

GIANPIETRO MORAS[1], RATHIN CHOUDHURY[2], JAMES R. KERMODE[2], GABOR CSÁNYI[4], MICHAEL C. PAYNE[3], AND ALESSANDRO DE VITA[2]

[1] *Institut für Zuverlässigkeit von Bauteilen und Systemen, University of Karlsruhe, Karlsruhe, Germany; Fraunhofer Institut für Werkstoffmechanik, Wöhlerstrasse 11, 79108 Freiburg, Germany, e-mail: moras@iwm.fraunhofer.de*

[2] *Department of Physics, King's College London, London WC2R 2LS, UK, e-mail: rathin.choudhury@kcl.ac.uk; alessandro.de_vita@cl.ac.uk; james.kermode@kcl.ac.uk*

[3] *Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, Cambridge, CB3 0HE, Cambridge, UK, e-mail: jrk33@cam.ac.uk; mcpl@cam.ac.uk*

[4] *Engineering Laboratory, University of Cambridge, CB2 1PZ, Cambridge, UK, e-mail: gc121@cam.ac.uk*

Abstract: The atomistic simulation of many processes in materials involves large-size model systems where different levels of complexity need to be described simultaneously. While accurate quantum mechanical simulations of large-size systems are usually not affordable, less computationally intensive classical models are not suitable for the description of many chemical processes. Hybrid (quantum/classical) modelling schemes are required in these circumstances. Here, we describe the "Learn on the fly" (LOTF) hybrid molecular dynamics scheme. Some technical aspects of this technique are illustrated through a series of examples of its applications to multiscale processes in silicon

Keywords: Quantum/classical atomistics, Hybrid modeling, Multiscale computations

1.1. INTRODUCTION

Molecular dynamics (MD) plays a very important role in the study of many kinds of physical, and more recently, biological processes, gaining insight at the atomistic level which in turn complements or can drive experimental work. Progress in the field has been driven by the two pronged advance of hardware and software, allowing the treatment of successively larger systems over successively longer times. However, we are far from an ideal world in which any and every problem can be studied with ab initio molecular dynamics, the problem domain being currently

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 1–23.
DOI 10.1007/978-1-4020-9785-0_1,

restricted to a few hundred atoms for tens of picoseconds at most, due to the limits of current technology. By loosening the accuracy of the quantum description to, for example, tight-binding, the domain can be further extended to thousands of atoms for longer times, but still there remain problems for which this is insufficient and quantum-mechanical techniques have to be replaced by simpler, but less transferable, classical interatomic potentials.

Within this *hierarchical multiscale* approach [1], physical quantities obtained by simulations performed at one length (or time) scale (e.g. ab initio MD) are used to parametrise models which are, in turn, used at a larger scale (e.g. classical interatomic potentials). However, many physical processes involve different, strongly coupled, length scales that must be studied *simultaneously*. This often happens when the influence of macroscopic forces produces structural or chemical changes at a much shorter-ranged, local level. Important examples of such *multiscale* phenomena in material science are stress-induced defect processes (e.g. diffusion of point defects, motion of dislocations, brittle fracture). For example, the propagation of a crack tip through a brittle material, or the motion of dislocations within a bulk matrix, involve the breaking and rebonding of atoms bordering the crack front or the dislocation core, respectively, while simpler stress concentration phenomena take place in a much larger surrounding area. In cases such as these the use of a uniform-accuracy Hamiltonian to model the entire system is not a viable strategy. A fully quantum-mechanical description of the model system is usually unfeasible, as well as wasteful since a simpler classical model would be sufficiently accurate to describe most of the system. However, less computationally intensive classical models are not suitable for the description of the chemically complex region of the system. Therefore, hybrid quantum/classical modelling schemes are required in these circumstances.

Several multi-Hamiltonian approaches have been proposed where the atomistic model system is divided into a chemically relevant quantum-mechanical (QM) region and a larger region which is described by a simpler classical interatomic potential [2–4]. The most serious difficulty when constructing such hybrid schemes is finding an effective and generally applicable treatment of the boundary between the two regions. Here, we describe a method which allows the boundary problem to be tackled in the context of hybrid quantum/classical molecular dynamics: the "Learn on the fly" (LOTF) technique [5–9]. Some technical aspects of this hybrid scheme are described through examples of its applications to multiscale processes in silicon, such as the motion of dislocations, brittle fracture and the growth of hydrogen-induced platelets.

1.2. THE LOTF SCHEME

1.2.1. Reconciling the Boundary

As outlined in the introduction, the main difficulty when developing hybrid quantum/classical atomistic schemes is finding an effective description of the boundary between the quantum-mechanical (QM) and the classical regions in which the

system is divided. Within standard quantum mechanical/molecular mechanical (QM/MM) schemes developed for biological systems [2, 3], the total energy of the hybrid system is obtained as the sum of the quantum-mechanical energy of the QM region, the classical energy of the rest of the system, and a term representing the interaction between the two subsystems. This requires the development of complex and dedicated techniques for the boundary treatment which usually prevent atoms from entering or leaving the QM region during MD simulations. As a result, the QM/MM simulation of many dynamical multi-scale processes in materials, such as the motion of dislocations or the propagation of brittle fracture, is technically very difficult. However, allowing the QM region to move during the MD simulation following the chemically complex region of the system (i.e. the dislocation core or the crack tip) is both necessary and indeed the only way to keep the number of atoms selected for the QM treatment low. Finally, very few attempts have been made so far to combine more than two Hamiltonians (see e.g. [10]), since the boundary problem is such cases is even more complex.

This problem can be solved by following an alternative route. Firstly, we avoid the complex task of deriving a combined Hamiltonian from the separate quantum-mechanical and classical Hamiltonians, and instead we focus on local quantities, such as the forces acting on the atoms, which can still be used to perform MD simulations. Secondly, we want the forces at each time step of the MD to be the derivatives of a *unique* Hamiltonian: this is a main idea behind the LOTF scheme and is obtained in the following way. We choose a unique classical Hamiltonian, and at each time step we adjust its parameters to reproduce accurate quantum-mechanical forces in those regions of the systerm where the accuracy of a classical potential would not be sufficient to describe the local physical processes (e.g. at the advancing crack tip). In general this unique Hamiltonian is complicated, and we have found that it is not always sufficient to simply vary the parameters of a pre-existing classical potential; this is discussed in more detail in Section 1.2.3 below.

Within the LOTF scheme, the boundary problem is therefore tackled in two separate steps:

Evaluation of the QM forces. Accurate QM forces are calculated during the simulation only in those regions of the system where the accuracy of the classical interatomic potential needs to be improved (i.e in the QM region of the system). If any of the atoms flagged for the QM treatment — i.e., those for which the QM forces are used for MD — were located on the surface of the zone treated by QM in the calculation (e.g., created by simply carving out the QM region from the system), large errors would be introduced by the presence of the artificial surface so introduced. This problem can be avoided by including in the QM calculation a "buffer zone" of finite width surrounding the QM-flagged set of atoms, and then retaining for use in the MD the forces on the "internal" QM-flagged atoms only (cf. Section 1.2.2).

Force matching. Forces coming from different Hamiltonians may suffer from an inconsistency across the boundary between the two regions. For example, force components acting along a bond crossing the boundary do not

necessarily obey an "action-reaction" principle. Tuning the parameters of a unique Hamiltonian to reproduce all the forces on the atoms avoids this inconsistency (any inconsistency in the forces across the boundary is smoothed out in the fit procedure) automatically enforcing the action-reaction principle.

A detailed description of the LOTF scheme is provided in the following parts of this section.

1.2.2. Evaluation of the QM Forces

At each time step of the LOTF MD, all the regions of the system which are not suitably described by the classical model are selected for the QM treatment. As already mentioned, this continuous selection procedure allows for small, mobile QM regions to be considered at each time step. However, it also requires a robust algorithm for selecting the QM region. The selection criteria we use are typically based on geometric and topological information (e.g. atom coordination, bond lengths, bond angles) and are defined by the user in order to track the regions of the system which are important for the particular physical process under investigation. Some details regarding the QM selection algorithms applied to the motion of dislocations and the propagation of brittle fracture in silicon will be discussed in Section 1.3.

Once the atoms in need of a quantum treatment are identified, the chosen quantum-mechanical "black box" engine (e.g. a Density-Functional-based scheme or a Tight Binding Hamiltonian) is used to compute the forces acting on the atoms of the QM region. As explained above, in order to avoid any errors originating from spurious QM "surface" atoms, we carve out a larger subsystem containing the desired QM region and a surrounding buffer zone. The QM blackbox is used to perform the calculation of the forces for this entire enlarged subsystem. Only the forces on the atoms composing the original QM region are retained while the forces on the atoms of the buffer zone, which are affected by the presence of the artificial surfaces, are discarded (i.e. not used for augmenting the classical model). Tests show that a $\sim$1 nm wide buffer zone allows very accurate force calculation in silicon systems (force component errors with respect to the fully QM calculation on the whole test system are typically lower than 0.01 eV/Å) [7]. The size of the buffer zone can be further reduced by providing a suitable termination of the dangling bonds at the outer cluster surface, using for example hydrogen atoms in a silicon system.

Based on the same idea, an alternative force evaluation scheme can be used where the forces on each atom in the QM region are computed independently. For each atom in the QM region, a spherical *cluster* is carved out, centred on the given atom. The QM forces are then calculated and only the force on the central atom of each cluster is retained. In this case, the radius of the spherical cluster is equivalent to the size of the aforementioned buffer zone and chemical termination of the cluster surfaces can be used to reduce the cluster radius (or to increase the accuracy of the force calculation for a given radius). Since each cluster calculation is independent, this scheme allows a very efficient parallelisation of the force evaluation, requiring minimal bandwitdth from the inter-processor communication hardware [7].

1.2.3. Force Matching

The QM forces calculated at each time step of the LOTF MD in a selected region, are used to locally augment the classical potential. This is achieved by means of a force matching procedure. The simplest force matching approach is to modify an existing classical potential. A suitably chosen subset of the classical Hamiltonian's parameters is let free to take different values across the system (e.g. each two-body term of the classical potential has an independent set of parameters). These parameters are then adjusted to reproduce the QM forces through the minimisation of the objective functional

$$F(\{\alpha\}) = \sum_{i=1}^{N} |\mathbf{F}_i^{\text{target}} - \mathbf{F}_i(\{\alpha\})|^2, \tag{1-1}$$

where $\{\alpha\}$ is the subset of varying parameters, $\boldsymbol{N}$ is the total number of atoms and

$$\mathbf{F}_i^{\text{target}} = \begin{cases} \mathbf{F}_i^{\text{QM}} & \text{if atom } i \in \text{ QM region} \\ \mathbf{F}_i^{\text{classical}} & \text{if atom } i \notin \text{ QM region.} \end{cases} \tag{1-2}$$

The parameter optimisation does not need to take place throughout the whole system. We can often just adjust the parameters of the classical Hamiltonian in a fitting region including the QM zone and a surrounding "crust" region as far as the crust is thick enough to allow a smooth transition and small errors in the force matching. Tests performed on Si and Si/H systems show that a fitting region including the QM zone and a surrounding nanometer-sized region is sufficient to obtain accurate force matching (i.e. force matching errors smaller than 0.01 eV/Å) using a few tens of conjugate gradient steps in parameters space.

This force matching approach has been successfully used to perform LOTF simulations on silicon systems, where the Stillinger-Weber (SW) potential [11] was used as the classical Hamiltonian [6, 7]. Nevertheless, adjusting the parameters of the classical Hamiltonian has some drawbacks:

1. Minimizing the objective functional of Eq. (1-1) requires the derivatives of the objective functional with respect to the parameters to be evaluated at each step of the minimisation. Expressions for these derivatives have thus to be coded for each different force field used to describe the system.
2. Often, no classical potential is available which is able to provide a reasonably accurate description of the system. This means that the difference between the classical and the QM forces in the QM region can be large and therefore a slight modification (a parameter variation within a reasonably small domain) of the classical potential might not be sufficient to achieve an accurate force matching.

For these reasons, we have chosen to extend the unique Hamiltonian used for force matching: instead of modifying the parameters of the classical potential, a

general *adjustable potential* (AP) can be used to reproduce the differences between the classical and the QM forces. This suitably parametrised potential is then added to the classical potential and the forces used in the MD are derived from the total potential. In practice, both the classical and the quantum-mechanical forces are calculated as described in Section 1.2.2. Then, the parameters $\{\alpha\}$ of the AP, which also in this case are free to take different values across the system, are tuned in order to minimise the objective functional

$$F(\{\alpha\}) = \sum_{i=1}^{N} |(\mathbf{F}_i^{\text{target}} - \mathbf{F}_i^{\text{classical}}) - \mathbf{F}(\{\alpha\})_i^{\text{adj}}|^2, \tag{1-3}$$

where $\mathbf{F}(\{\alpha\})_i^{\text{adj}}$ are the forces calculated as gradients of the AP and

$$\mathbf{F}_i^{\text{target}} = \begin{cases} \mathbf{F}_i^{\text{QM}} & \text{if atom } i \in \text{QM region} \\ \mathbf{F}_i^{\text{classical}} & \text{if atom } i \notin \text{QM region.} \end{cases} \tag{1-4}$$

Therefore, after a successful force matching, we have, for each atom,

$$\mathbf{F}^{\text{target}} \simeq \mathbf{F}^{\text{classical}} + \mathbf{F}^{\text{adj}}. \tag{1-5}$$

We can therefore propagate the MD by using the potential

$$V(\mathbf{r},\{\alpha\}) = V^{\text{classical}}(\mathbf{r}) + V^{\text{adj}}(\mathbf{r},\{\alpha\}), \tag{1-6}$$

the force on each atom being

$$\mathbf{F} = -\nabla_{\mathbf{r}} V = \mathbf{F}^{\text{classical}} + \mathbf{F}^{\text{adj}}. \tag{1-7}$$

Within the *adjustable potential* approach, the force matching procedure is independent of the classical Hamiltonian used to describe the system. Therefore, the derivatives of the AP with respect to the parameters can be coded once and for all. Moreover, both classical and quantum-mechanical forces are obtained from *black box* engines, which take atomic coordinates and lattice parameters as input, returning the force on each atom. This increases the flexibility of the scheme, since new force models can be added quickly and easily. A fully flexible adjustable potential can, furthermore, be constructed so that it guarantees accurate forces even for local system configurations for which the underlying classical potential used is far from accurate.

1.2.3.1. The Adjustable Potential

We want the choice of the AP analytical form to be independent of the particular system under investigation, and of the classical and quantum-mechanical Hamiltonians. Therefore, the AP should offer a good compromise between expressive power and robustness. Moreover, since the AP parameters are adjusted on the fly, the AP should

change smoothly as the parameters change. Finally, the AP derivatives with respect to its parameters, which need to be computed at each step of the force matching procedure, should be ideally easy to code and fast to evaluate [9, 12]. According to these criteria, a simple linear spring adjustable potential has been formulated, using bond lengths r_{ij} and bond angles θ_{jik} as fundamental coordinates. The linear spring potential can be written as a sum of two-body and three-body terms

$$V_{\text{spring}}(\mathbf{r},\{\alpha\}) = \sum_{\substack{i,j \\ i<j}} \alpha_{ij} r_{ij} + \sum_{\substack{i,j,k \\ i<j<k}} \beta_{jik} \cos\theta_{jik}, \tag{1-8}$$

where $\{\alpha\}$ is the set of free parameters α_{ij} and β_{jik} and both the two-body and three-body terms are linear in the interatomic distance r_{ij} and in $\cos\theta_{jik}$, respectively. An important consequence of using a linear potential is that the objective function in Eq. (1-3) can be minimised by singular value decomposition (SVD), thus allowing fast linear algebra routines to be used. This analytical form has been tested on Si, Si/H and C systems [13, 14]. Very accurate force matching (i.e. force matching errors smaller than 0.01 eV/Å) can be achieved by using the two-body terms only and a nanometer-sized buffer region for the fitting procedure.

1.2.4. The LOTF Predictor-Corrector Scheme

The parameters of the classical potential (or Adjustable Potential) do not generally need to be adjusted at each time step of a LOTF MD simulation. In fact, if the fractional variation of the AP parameters is sufficiently small along the trajectory, the AP can be used with unchanged parameters for a small number of time steps after each force matching procedure. This allows a significant speed up of the MD simulation, since the computationally expensive evaluation of the QM forces is only carried out when the AP parameters actually need to be adjusted. In practice, the LOTF MD simulation can be performed by using a *predictor-corrector* scheme, as follows (a diagram of this procedure is shown in Figure 1-1). Starting from a point $\mathbf{R}_0$ in phase space, the QM forces are calculated on the selected atoms and the parameters of the adjustable Hamiltonian are optimised in order to reproduce these accurate forces. After the parameter optimisation, which gives AP parameters $\{\alpha_0\}$, a small number $\boldsymbol{N}$ of MD time steps are performed using this fixed set of parameters to arrive at point $\mathbf{R}_1'$ (this part is referred to as *predictor* or *extrapolation* part). Here, a new QM force evaluation is performed in the QM region and the AP parameters are re-optimised. Simply continuing the MD from this point using the new set of parameters $\{\alpha_1\}$ would not make use of the knowledge of the local optimal parameter variation occurred between the last two QM force evaluations, and would furthermore lead to a spurious discontinuity in the second derivative of the trajectories. Therefore, we return to the initial atomic positions, $\mathbf{R}_0$, and redo the MD for $\boldsymbol{N}$ time steps using a linear interpolation of the parameters between $\{\alpha_0\}$ and $\{\alpha_1\}$ (*corrector/interpolation* part). If t_0 is the time corresponding to the configuration $(\mathbf{R}_0,\{\alpha_0\})$ and $t_1 = t_0 + N\Delta t$, where Δt is the time step of the MD,

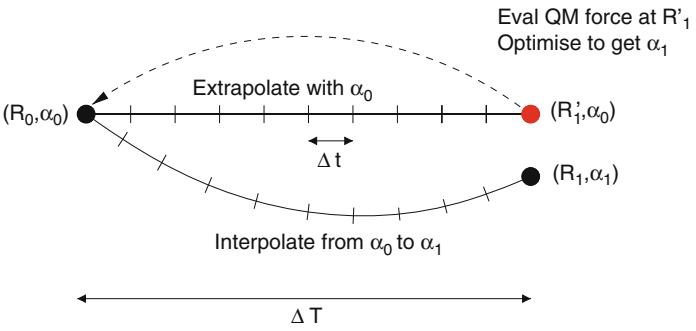


Figure 1-1. The LOTF predictor-corrector scheme. The MD is propagated from $\mathbf{R}_0$ using fixed parameters $\{\alpha_0\}$ for a small number of time steps N (predictor/extrapolation). New QM forces are calculated in the QM region after the predictor cycle ($\mathbf{R}'_1$) and the force matching procedure is performed giving a new set of parameters $\{\alpha_1\}$. The MD is then redone for N time steps starting from $\mathbf{R}_0$ and using parameters linearly interpolated between $\{\alpha_0\}$ and $\{\alpha_1\}$ (corrector/iterpolation)

then the set of parameters used at $t_0 + m\Delta t\,(0 \leq m \leq N)$ during the corrector cycle is calculated as

$$\{\alpha_m\} = \{\alpha_0\} + \frac{m}{N}(\{\alpha_1\} - \{\alpha_0\}). \tag{1-9}$$

At the end of the corrector cycle we arrive at a point in phase space, $\mathbf{R}_1$, which is different from $\mathbf{R}'_1$. However, if N is suitably small, this point is within the phase space region where the parameters $\{\alpha_1\}$, optimised in $\mathbf{R}'_1$, provide very accurate forces, and the dynamics can be continued from ($\mathbf{R}_1$,$\{\alpha_1\}$) iterating this predictor-corrector cycle. Note that using a smaller time step would correspond to an increase of N. However, the MD time step is already the maximum allowed by the system's physical conditions. Therefore, using a time step N times larger and no predictor-corrector scheme would compromise the stability of the simulation so that the predictor-corrector procedure is associated to a genuine overall speed up.

The robustness of the AP form determines the choice of the number of extrapolation/interpolation time steps, N. Tests performed tuning the parameters of the SW potential or of a two-body linear spring potential, and using different QM black boxes, show that $N = 5$–10 time steps can be used to accurately model silicon systems at high temperatures. Although the proposed adjustable potentials are equally accurate in reproducing QM forces, we might expect the interpolation behaviour of the simpler spring potential to differ from that of the Stillinger-Weber based approach. This has been tested by comparing the forces given by the LOTF adjustable Hamiltonian during the extrapolation and interpolation cycles ($\mathbf{F}_{\mathrm{LOTF}} = \mathbf{F}_{\mathrm{classical}} + \mathbf{F}_{\mathrm{AP}}$) with the full hybrid force (i.e. the forces calculated by means of the QM blackbox within the QM zone, and the classical Hamiltonian outside the QM region) calculated at each time step of the predictor and corrector parts. The error

$$\mathbf{F}_{\text{error}} = \mathbf{F}_{\{\text{QM,classical}\}} - \mathbf{F}_{\text{LOTF}} \tag{1-10}$$

gives a measure of the transferability of the spring potential during the predictor-corrector loop. The test was carried out for a 64 atom silicon system at a temperature of 2000 K, with a MD time step $\Delta t = 1$ fs. The linear spring adjustable potential with two body springs only was used and the QM force model was the tight binding potential of Kwon et al. [15]. 1000 independent trajectories were generated by randomising the initial atomic positions and velocities. The force errors during one predictor-corrector cycle were recorded and then averaged over all the trajectories to produce distributions of RMS and maximum force errors as a function of time during the extrapolation and interpolation, as illustrated in Figure 1-2 for 10-step predictor-corrector loops. The force errors rise approximately linearly during the extrapolation part of the cycle, as we move away from the point in phase space (i.e., the set of atomic positions) where the potential parameters were fitted. Providing the range of validity of the new parameters fitted at the end of the extrapolation is large enough, the force errors should remain small throughout the interpolation. We must choose the number of interpolation steps N appropriately to ensure this

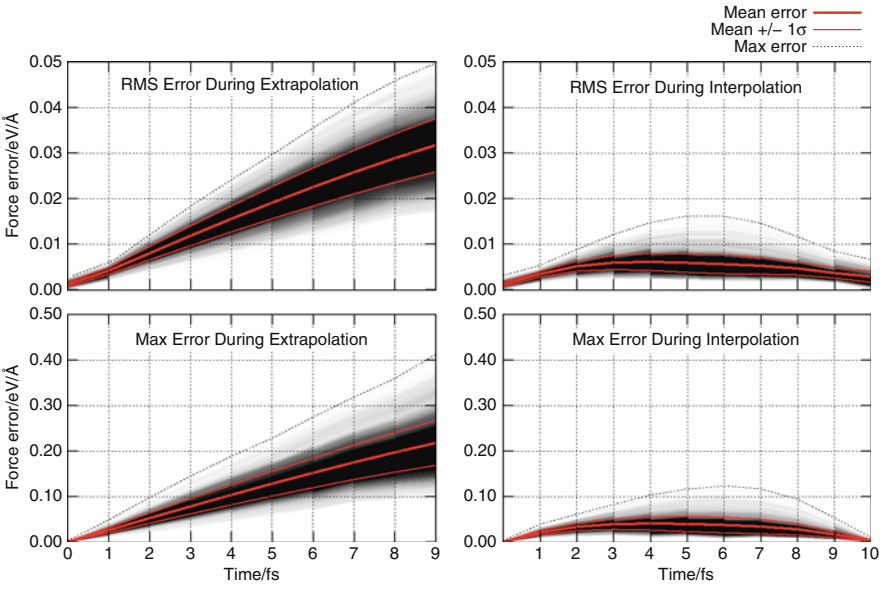


Figure 1-2. Transferability of the linear spring potential. The system used to test the adjustable potential was a 64 atom silicon bulk cube with 5 QM atoms, at a temperature of 2000 K, with MD timestep $\Delta t = 1$ fs and $N = 10$. Force errors during one predictor-corrector cycle were averaged over many independent trajectories. The *gray-scale* colouring shows the density of force errors as a function of time, with *black* corresponding to high and *white* to low densities. The mean and standard deviation of the distribution are indicated by the *red lines*, and the extreme values by the *black dashed lines*. The linear spring adjustable potential with two body springs only was used. QM force model is the tight binding Hamiltonian of Kwon et al. [15]

is the case. The appropriate scale is set by the accuracy to which the forces have been computed; for a DFT-based self-consistent QM Hamiltonian (CASTEP [16]) this is normally around 0.05 eV/Å. We can see from Figure 1-2 that $N = 10$ gives acceptable accuracy: the RMS deviation typically remains smaller than 0.01 eV/Å, and the maximum force error is typically below 0.1 eV/Å. The transferability tests were repeated with a variety of different adjustable potentials, including the original Stillinger-Weber potential with variable parameters and a form of the adjustable potential using cubic splines instead of linear springs. There was no significant difference in the distribution of force errors produced by the different potentials. Moreover, the force error distribution is essentially unchanged when three body springs are included. The choice of linear or cubic parameter interpolation makes only a very small difference to the measured force errors, as does repeating the predictor corrector cycle two or more times.

1.3. SELECTION OF THE QM REGION: AN HYSTERETIC ALGORITHM

As introduced in Section 1.2.2, the selection of the quantum region has to be implemented using geometric and topological criteria. However, the instantaneous atomic positions cannot always be used for this purpose, since they often fluctuate too rapidly to allow the creation of a stable QM region. Unnecessary rapid oscillation of the set of atoms forming the QM region should be avoided since they may cause a number of problems. Atoms that move in and out of the QM region can generate spurious heat as they do so, since the two energy models are trying to drive the system towards different configurations. Most importantly, from a practical point of view, such changes in the number of QM atoms are inefficient since ab initio calculations can be sped up very significantly by reusing the electronic density and wavefunctions from the previous timestep.

For these reasons, it is useful to use time-averaged atomic positions for the QM region selection, in order to filter out the fast optical phonons. The time-averaged coordinates are defined as

$$\tilde{\mathbf{r}}(t) = (1 - e^{-\Delta t/\tau}) \sum_{n=0}^{\infty} e^{-n\Delta t/\tau} \mathbf{r}(t - n\Delta t), \tag{1-11}$$

where Δt is the MD time step and τ the averaging time constant [7]. However, in many circumstances this does not allow us to completely avoid undesirable fluctuations. Difficulties occur in situations where it is impossible to choose the averaging time τ to be large enough to exclude atomic oscillations and at the same time small enough to still capture rapid bond breaking events. An important example is the case of a fast-moving crack: it has been found that τ needs to be below about 100 fs to correctly follow a moving crack in silicon, but this is of the same order as the thermal oscillation period of a single Si–Si bond [13]. Therefore, atoms at the edge of the QM region tend to pop in and out as they vibrate backwards and forwards. Another example where using time-averaged atomic positions is not sufficient to

guarantee a stable QM region, is the motion of dislocations. Atoms at a dislocation core may pass through bulk like configurations as they are being modelled, for long enough timescales that they are no longer identified as needing quantum treatment. To solve this problem, a selection algorithm with hysteresis has been developed [13]. Essentially, once an atom has been flagged for quantum treatment through topological considerations based on its time-averaged position, two spheres are defined around it, one inner and one outer with radii R_{in} and R_{out} respectively. An atom within a distance R_{in} from the flagged atom is also flagged for quantum treatment, extending the QM zone. However, it is not deselected until it travels further than R_{out} from the central (flagged) atom, so that entry is effectively easier than escape. Modelling the QM zone in this way suppresses unwanted fluctuations, and provides more reliable tracking of the QM atoms. The dislocation and the brittle fracture studies mentioned above are discussed in the next parts of this section, together with some details of the hysteretic selection algorithm applied during LOTF simulations of these two systems.

1.3.1. A Screw Dislocation Study

Understanding the properties of dislocations in silicon has important implications, due to their electrical properties and effects on devices, acting as trapping or scattering centres for charge carriers. Much work has been done on dislocations in silicon, both in theory and experiment, elucidating the many types with their different geometries and core structures. Previously, ab initio studies have been carried out to obtain a picture of the different types of kink that form from the dislocation cores, estimating the energy of formation and thus deducing the most likely mechanisms for kink migration along the dislocation line [17]. In a high Peierls' barrier material such as silicon, dislocation glide occurs through the formation and propagation of kinks rather than via coherent motion of the line itself. Thus far computational studies have been restricted to focus on the static properties of such dislocations, and the dynamics have been modelled mainly through kinetic monte-carlo techniques [18]. Using the LOTF method, however, it is possible to perform a MD study of the dislocation glide, a problem which naturally fits the LOTF paradigm, since the majority of the system remains in a bulk configuration, conferring strain onto the moving dislocation core, where the localised bond breaking/forming events take place. In this section we give a brief introduction to the screw dislocation in silicon.

The most important mobile dislocations in silicon are the screw and the 60° dislocations. Our current work focusses on motion of the screw dislocation at high temperature in the presence of grain boundaries [19, 20]. Above 600 K, the perfect screw dislocation dissociates into two 30° partial dislocations bounded by a stacking-fault (SF) region. The structure of the perfect screw has been studied using ab initio techniques [21], as has that of the 30° partial [22]. Figure 1-3 shows the core structure of two 30° partials, separated by a SF region. In fact, the 30° partial is thought to govern the dislocation dynamics in silicon. The atoms along the core are dimerised and the dimers must be broken in order to form the kinks which are the precursor of migration, with a high energetic cost [17].

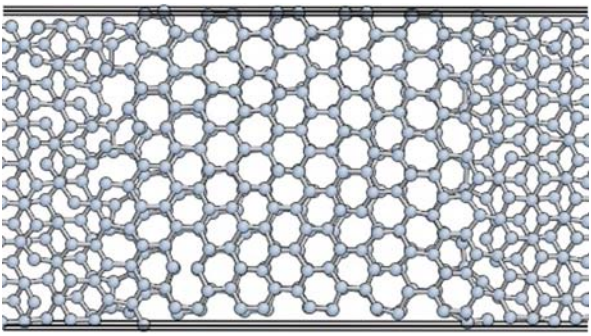


Figure 1-3. Atomic structure of the dissociated screw dislocation in silicon, separated into two 30° partials which are bounded by a SF

In order to model the dislocation glide, we must be able to track the moving dislocation core, which forms the QM zone of our LOTF calculation. Firstly, we must identify the atoms which are undergoing bond-breaking processes. This can be done simply by monitoring changes in the time-averaged number of nearest neighbours for atoms in the dislocation glide plane. We have seen that atoms in the unperturbed core are dimerised in the [110] direction, so these are also flagged for quantum treatment. However, these criteria alone are not enough, as the dynamics is performed at high temperature, around 1400 K. As a consequence of the high temperature there are large fluctuations of the atomic positions so that atoms which should be flagged for quantum treatment may be overlooked. Usage of the hysteretic selection of the QM region overcomes this problem of large fluctuations, as can be seen from Figure 1-4. The different coloured lines correspond to different choices of inner and outer radii in the definition of the hysteretic quantum zone. The black line, corresponding to a 3 Å inner radius and a 5 Å outer radius, drops by half after 50 ps of simulated time, because one of the partial cores has been "lost" by the hysteretic algorithm. In practice, having inner and outer radii which are too small reduces the effectiveness of the hysteretic quantum zone scheme which proves not robust enough to identify the fluctuating dimers, and the observed drop is due to the fact that one remaining core only is being tracked. With a more robust choice of inner radius of 4 Å the quantum selection remains broadly stable throughout the run (blue and red lines), despite the different outer radii (7 Å for the red line and 10 Å for the blue).

1.3.2. Brittle Fracture

One of the best known examples of a strongly coupled multiscale system is the fracture of brittle materials. The conditions for crack propagation are created by stress concentration at the crack tip, and depend on macroscopic parameters such as the loading geometry and dimensions of the specimen [23–26]. In real materials, however, the detailed crack propagation dynamics are entirely determined by

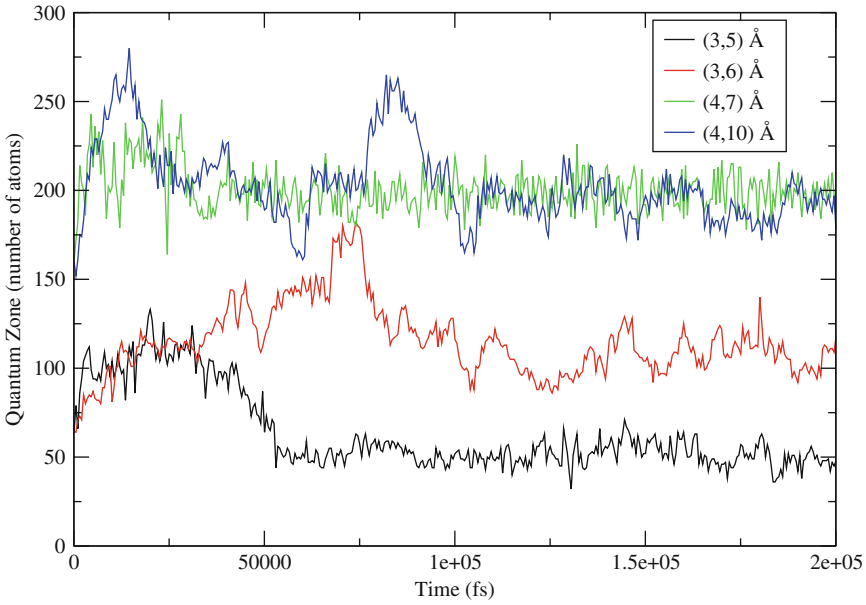


Figure 1-4. Effect of using hysteretic QM zone in tracking the migrating dislocation core. The pair of numbers corresponding to each of the *colours* in the picture label are the radii of the *inner* and *outer circles*, respectively

atomic scale phenomena since brittle crack tips are atomically sharp and propagate by breaking bonds, one at a time, at each point along the crack front [27,28]. This means the tip region is primarily a one dimensional line, perpendicular to the direction of propagation, and so it should be possible to define a contiguous embedding region to be treated with a more accurate model in a hybrid simulation. There is a constant interplay between the length scales because the opening crack gives rise to a stress field with a singularity at the tip [29], as illustrated in Figure 1-5, and in turn it is this singular stress field which breaks the bonds that advance the crack. Only by including the tens of thousands of atoms that contribute significantly to the elastic relaxation of this stress field can we hope to accurately model the fracture system, and thus a multiscale approach is essential.

The LOTF method is ideally suited to studying brittle fracture since the use of a buffer zone to yield accurate quantum forces allows the the quantum region to be made small and mobile. This requires a robust selection algorithm to follow the crack tip as it moves and identify the atoms that need to be treated with quantum mechanical accuracy. This is a difficult problem since the timescales of thermal vibration and crack motion are not well separated. The hysteretic selection algorithm described above provides an effective solution to this problem.

We flag atoms as "active" when they change their bonding topology, and then construct embedding ellipses around each active atom. The set of active atoms is seeded with a few atoms near to the crack tip at the start of the simulation. Figure 1-6

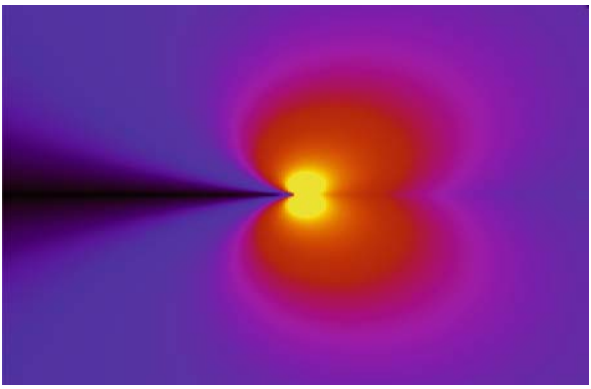


Figure 1-5. Maximum principal stress near the the tip of a crack under uniaxial tension in the opening mode, from the linear elastic solution. *Black* areas are the least stressed and *yellow* the most

illustrates how the algorithm works in a simple case with only two active atoms — in reality there could be several hundred. Ellipses with different radii are used to define inner and outer selection regions, and then the hysteretic algorithm ensures that atoms near the edges of the QM region do not oscillate in and out of the active region. Elongated ellipses allow the embedding region to be biased forwards so that the QM region always extends ahead of the crack tip.

As the crack moves on, we can stop treating atoms behind the crack tip quantum mechanically. We cap the size of the QM region at some limit N_{quantum} based on our computational capability — this can be several hundred atoms for a tight binding simulation, or of the order of a hundred for an ab initio simulation. By keeping track of the order in which atoms became active, we can remove them from the QM region in a consistent fashion. An additional condition prevents atoms further than a threshold distance away from the centre of mass of the current QM region from becoming active.

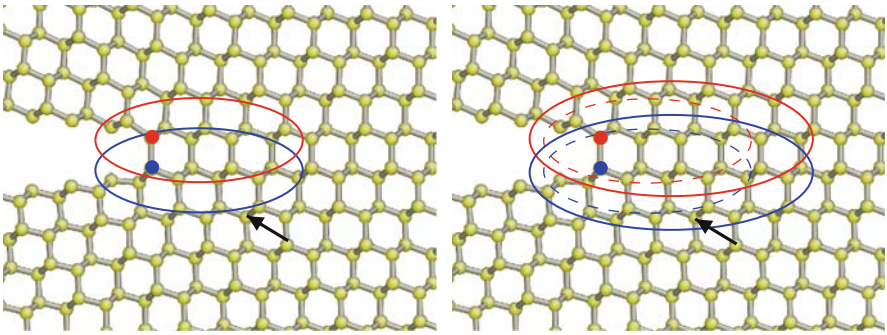
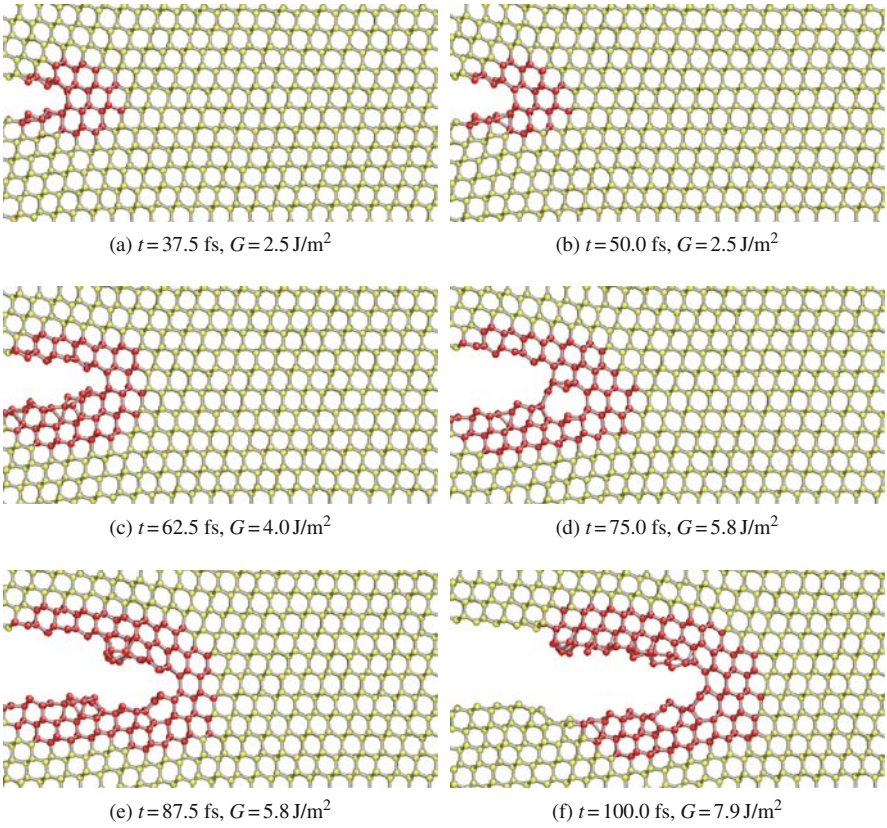


Figure 1-6. Hysteretic QM selection algorithm applied to crack tip region. The *red* and *blue atoms* are considered "active", and are used to define inner (left panel) and outer (right panel) selection regions. The atom indicated with the black arrow remains selected despite oscillating in and out of the inner region providing that it stays inside the outer region



(a) $t = 37.5$ fs, $G = 2.5\,\mathrm{J/m^2}$ (b) $t = 50.0$ fs, $G = 2.5\,\mathrm{J/m^2}$

(c) $t = 62.5$ fs, $G = 4.0\,\mathrm{J/m^2}$ (d) $t = 75.0$ fs, $G = 5.8\,\mathrm{J/m^2}$

(e) $t = 87.5$ fs, $G = 5.8\,\mathrm{J/m^2}$ (f) $t = 100.0$ fs, $G = 7.9\,\mathrm{J/m^2}$

Figure 1-7. Snapshots from LOTF tight binding MD simulation at a temperature of 300 K using the tight binding model of Kwon et al. [15]. The *red atoms* are treated quantum mechanically and the *yellow atoms* with the classical potential. Atomistic visualisations were produced with AtomEye [32]

Hybrid fracture simulations on the silicon (111) cleavage plane using any of variety of tight binding potentials [15,30] or the CASTEP [16] density functional code near the crack tip combined with a classical atomistic potential [11] further away yield brittle fracture. The calculations reveal a novel sub-critical crack tip reconstruction that plays an important role in the fracture process, full details of which are published elsewhere [31]. Figure 1-7 shows snapshots from a LOTF tight binding simulation of brittle fracture.

1.4. TOWARDS CHEMICAL COMPLEXITY: HYDROGEN-INDUCED PLATELETS IN SILICON

The investigation of the influence of impurities and corrosive agents on failure processes is a natural extension of the brittle fracture study. Environment-assisted cracking processes are a common cause of limited lifetime of engineering

components. Nevertheless, their complex underlying mechanisms remain poorly understood at the atomic level. Relevant examples are stress-corrosion-induced cracks at the Si/SiO_2 interface in electronic devices and hydrogen-embrittlement in metallic components.

In this section, we focus on a particular example of environment-assisted cracking of crystalline silicon which have been harnessed for the production of "silicon-on-insulator" devices in microelectronics through the "Smart Cut" technique [33]. This widely used technological process allows a thin layer of oxidized silicon to be transferred from a hydrogen-implanted silicon wafer onto a substrate after wafer bonding and heat-induced splitting. The latter is caused by the propagation of microcracks parallel to the implanted surface. Microcracks are thought to origin from the growth of hydrogen-induced platelets (HIPs) during the first stages of the high temperature annealing. HIPs are $\sim 10\,\text{nm}$ wide and $\sim 1\,\text{nm}$ thick disk-shaped extended defects which nucleate at room temperature after H-implantation. Most of these extended defects are located in the region where the final splitting takes place and lie on planes parallel to the implanted surface (i.e. the (100) planes for the technologically relevant Si(100) implanted wafers). Despite the important role played by HIPs in the Smart Cut process, a detailed atomic-level knowledge of both their nucleation and growth processes is still lacking. Both DFT-based [34, 35] and classical [36] simulation techniques have been used to study the structure and the formation mechanism of HIPs. The results of these simulations combined with those of experimental investigations [37] suggest that (100)-HIPs are composed of two facing internal H-terminated Si(100) surfaces. The surfaces are created either by substitution of Si–Si bonds with Si–H/H–Si structures or by coalescence of H-saturated silicon vacancies. However, using uniform accuracy to perform MD simulations of the thermally activated platelet growth is not a viable strategy. These simulations would require a combination of quantum-mechanical accuracy and large system sizes in order to couple the chemical reactions occurring at the platelet edges with the stress field in the surrounding Si crystal.

To solve this problem, we have applied the LOTF scheme to perform MD simulations on a realistically sized (100) HIP model system [38]. A $\sim 10\,\text{nm}$ wide (100) HIP is centered in a $\sim 35 \times 35\,\text{nm}^2$ crystalline silicon slab and periodic boundary conditions (PBCs) are applied perpendicularly to the slab plane. While a standard Stillinger-Weber interatomic potential is sufficient to describe silicon atoms in the crystalline phase, a QM technique is necessary to describe silicon and hydrogen atoms in the defective platelet region. During the LOTF MD simulation, the following atoms are selected for the QM treatment: (a) all hydrogen atoms; (b) all silicon atoms having at least one hydrogen atom among their first neighbours; (c) all under-coordinated Si atoms (defined as the Si atoms having less than 4 neighbours within a 2.5 Å distance, excluding those located on the outer slab surfaces); (d) the first neighbours of the atoms in (a), (b), and (c). Figure 1-8 shows a typical (100) HIP model system, where the hydrogen atoms and silicon atoms depicted in red are included in the QM region.

Within the LOTF predictor-corrector scheme, the Density Functional Tight Binding (DFTB) [39] formalism is used to calculate the forces on the atoms flagged

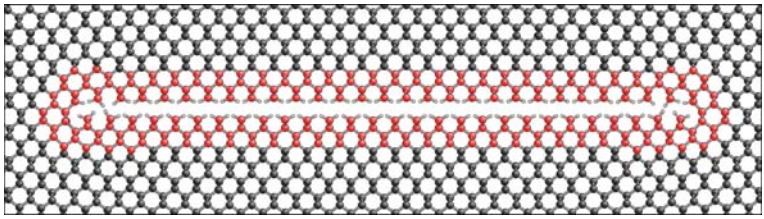


Figure 1-8. A realistic-sized (100)-HIP atomistic model: enlargement of the $\sim$ 10 nm wide HIP region. The whole system is a $35 \times 35\,\text{nm}^2$ silicon crystal. Periodic boundary conditions are applied along the direction perpendicular to the picture plane. The silicon atoms depicted in *red* and the hydrogen atoms are flagged for the QM treatment within the LOTF scheme

for the QM treatment whenever needed. An efficient calculation of the DFTB forces in this relatively large QM region (300-600 atoms, depending of the PBC orientation) is performed using the ("embarrassingly parallel") *multiple clusters* method described in Section 1.2.2. To determine the minimum size of H-terminated spherical clusters which allows accurate DFTB forces to be calculated in the QM region, we have tested the accuracy of the forces obtained through the *multiple clusters* method on a small model system. This is depicted in Figure 1-9a and is composed of two 2×1-reconstructed dihydride (100) Si surfaces. PBCs are applied on the plane parallel to the surfaces, while 12 Si planes are considered on the perpendicular direction. The DFTB forces calculated by using H-terminated multiple clusters on the H atoms and on the 4 outmost Si planes during a 400 fs MD at 800 K are compared to the forces calculated on the same atoms by using DFTB on the whole periodic system. In Figure 1-9b, the mean force component error for Si and H atoms is plotted as a function of the clusters radius. This test shows that a 7 Å cluster radius is sufficient to keep the error lower than 0.02 eV/Å for both Si and H atoms. Interestingly, the error obtained on the Si atoms using 8 Å clusters is observed to be larger than the error made by using 7 Å clusters, at the right end of the plot. The same behaviour has been observed in other periodic silicon systems [7]. A similar small peak of the error is observed when the diameter of the clusters becomes approximately equal to the size of the periodic cell. The cause of this slight temporary increase of the error is related to the simultaneous presence in the cluster of an atom and of its periodic image.

A two-body linear spring potential is used to match the differences between the hybrid (DFTB/SW) and the classical (SW) forces in the fitting region. The latter includes all the atoms in the QM region and four shells of first neighbours of the atoms at the boundary of the QM region. The maximum force component error in reproducing the target forces is again typically smaller than 0.02 eV/Å. This level of accuracy is obtained by setting to zero all the two- and three-body terms of the SW potential involving H atoms. Tests show that using a SW-like parametrisation for these terms give rise to higher force matching errors. This is due to the relatively poorer transferability of these parametrisations, sometimes causing the difference

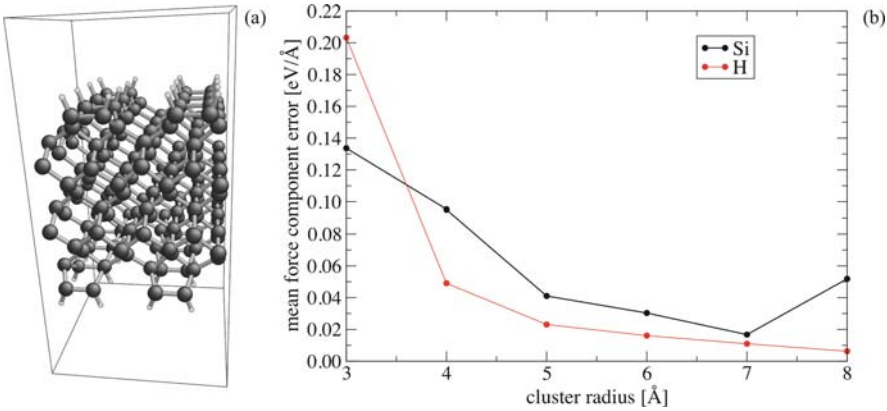


Figure 1-9. Accuracy of the QM force calculation using the multiple clusters approach. (**a**) Two 2 × 1-reconstructed dihydride (100) Si surfaces. PBC are applied on the plane parallel to the surfaces, while 12 Si planes are considered on the perpendicular direction. (**b**) The fully DFTB forces on the H atoms and on the 4 outmost Si planes during a 400 fs MD at 800 K are compared to the forces calculated using multiple clusters on the same atomic configurations. In this graph, the mean force component errors for Si and H atoms are plotted as a function of the clusters radius

between the classical and quantum-mechanical forces to be even larger than the quantum forces themselves.

Geometry optimization of the HIP model system has been performed through LOTF damped dynamics using a 10-step predictor-corrector loop. The transferability of the two-body spring potential during the predictor-corrector cycles, at a temperature of about 100 K, has been tested using the procedure described in Section 1.2.4 for the Si system case. A 1000 time-step MD was performed using 10-step predictor-corrector loops on two different HIP systems, using a 0.2 fs time step. The force error of Eq. (1-10) was calculated during each predictor-corrector loop. Figure 1-10 shows the evolution of the RMS and maximum force errors during the predictor and corrector cycles, averaged over all the predictor-corrector loops of the two simulations. While the error increases in a nearly linear way during the predictor part, it remains roughly constant during the corrector part where the time-averaged maximum force error is about 0.02 eV/Å.

1.4.1. The Atom-Resolved Stress Tensor

Besides being the starting point for subsequent finite-temperature LOTF MD, the relaxed atomic positions can be used to calculate the stress field distribution in the silicon crystal. This distribution does not suffer any discontinuity at the border between the classical and the QM regions since a unique Hamiltonian is used to calculate the forces on the whole system. Moreover, the parameters of the classical potential are slightly varied in order to reproduce as accurately as possible the elastic constants obtained by the QM blackbox. This guarantees an elastic matching between the two regions. As a final technical point, we now describe how the

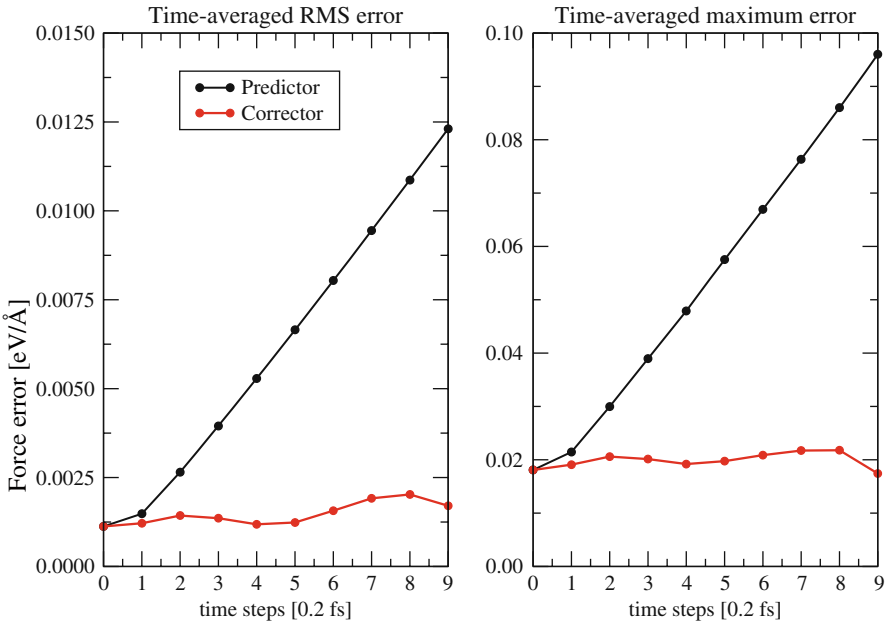


Figure 1-10. RMS (*left*) and maximum (*right*) force errors during 10-step predictor and corrector loops averaged during a 1000 time step LOTF MD. The simulations are performed on two different HIP systems at a temperature of about 100 K, using a two-body adjustable spring potential. The classical and QM Hamiltonians are the SW potential and the DFTB Hamiltonian, respectively

stress/strain tensor field can be practically evaluated at the atomistic scale and usefully compared with elasticity theory results. The atom-resolved values of the stress tensor are obtained from a local definition of the strain tensor through the linear theory of elasticity.

A definition of the atom-resolved strain tensor is obtained for all the four-fold coordinated Si atoms in the tetrahedral structure (all other atoms are assigned zero strain) by comparing the atomic positions to the atomic positions in the unstrained crystal. The four neighbours of an atom are used to build a local set of cubic axes. The components of the local deformation, with respect to the unstrained tetragonal structure, are combined into a matrix

$$U = \begin{pmatrix} | & | & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ | & | & | \end{pmatrix}, \tag{1-12}$$

where $\mathbf{e_i}$ ($i = 1,2,3$) is the component of the deformation along the i-th axes. This deformation can be decomposed into two contributions, one due to rotation and one due to strain. This is done by polar decomposition of the matrix U which gives $U = SR$, where R is a unitary matrix (rotation contribution) and S is a positive

semi-definite Hermitian matrix (strain contribution). If one considers the matrix $UU^{\dagger}$, where $U^{\dagger}$ is the Hermitian transpose of U, it turns out that

$$UU^{\dagger} = SS^{\dagger} \tag{1-13}$$

since $RR^{\dagger} = I$, where I is the identity matrix, R being a unitary matrix. Moreover, the eigenvalue decomposition of $UU^{\dagger}$ gives:

$$UU^{\dagger} = VLV^{\dagger}, \tag{1-14}$$

where V is the matrix of the eigenvectors of $UU^{\dagger}$ and L is the diagonal matrix of the $UU^{\dagger}$ eigenvalues. From Eqs. (1-13) and (1-14), we obtain

$$SS^{\dagger} = VLV^{\dagger} \tag{1-15}$$

and, being the eigenvalues of an Hermitian matrix real numbers,

$$S = VL^{\frac{1}{2}}V^{\dagger}. \tag{1-16}$$

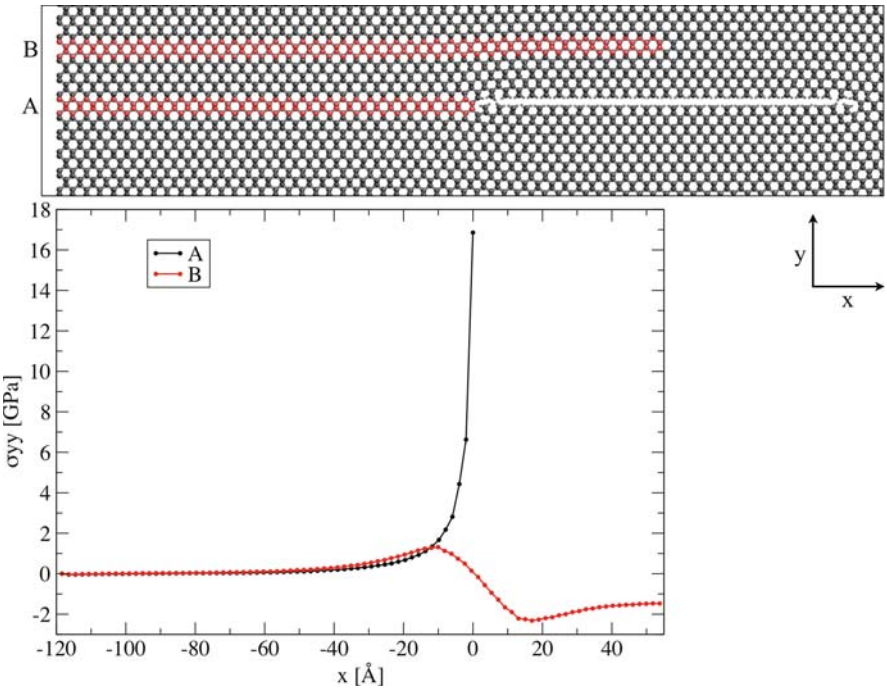


Figure 1-11. Atomic-resolved stress distributions: two cross-sections of the σ_{yy} component distribution are plotted as a function of the atomic x coordinate. The *black* and *red* curves refer to the regions A and B, respectively, whose atoms are depicted in *red*

Therefore S can be directly calculated from the eigenvalues and eigenvectors obtained by the diagonalisation of $UU^{\dagger}$. Once S has been calculated, R is obtained as $R = S^{-1}U$. The components of the strain tensor ε_{xx}, ε_{yy}, ε_{zz}, ε_{xy}, ε_{xz} and ε_{yz} can now be calculated by rotating S to align the local cubic axes with the reference Cartesian axes:

$$R^{\dagger}SR = I + \varepsilon = \begin{pmatrix} 1+\varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{xy} & 1+\varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{xz} & \varepsilon_{yz} & 1+\varepsilon_{zz} \end{pmatrix}. \tag{1-17}$$

The analysis of the stress distribution in our relaxed HIP systems shows that a large intensification of the tensile $\boldsymbol{\sigma}_{yy}$ component is observed close to the platelet edges, while a smaller compressive component occurs above and below the defect as a result of the two surfaces moving apart from each other during relaxation. Two cross sections of the $\boldsymbol{\sigma}_{yy}$ distributions are shown in Figure 1-11. Curves A and B refer to the two regions whose atoms are depicted in red. The curves are obtained by plotting the $\boldsymbol{\sigma}_{yy}$ component of the atom-resolved stress tensor in these regions as a function of the x coordinate.

We are currently investigating whether the large tensile stress intensification at the platelet edges could provide the driving force for HIPs growth during high temperature annealing.

ACKNOWLEDGMENTS

G.M. acknowledges the HPC-Europa programme for funding and computational resources and is grateful to Lucio Colombi Ciacchi for many insightful discussions and useful suggestions. J.R.K., G.C. and M.C.P. acknowledge support from the EPSRC portfolio grant EP/C523938/1. A.D.V. and R.C. acknowledge support from EPSRC grant EP/5C23938/1. G.C. acknowledges support from EPSRC grant EP/C52392X/1.

REFERENCES

1. Nieminen, R.M.: From atomistic simulation towards multiscale modelling of materials. J. Phys. Condens. Matter. **14**, 2859–2876 (2002)
2. Lin, H., Truhlar, D.G.: QM/MM: What have we learned, where are we, and where do we go from here? Theor. Chem. Acc. Theor. Comput. Model. Theor. Chim. Acta **117**, 185–199 (2007)
3. Ruiz-Lopez, M.F.: Combined QM/MM calculations in chemistry and biochemistry. J. Mol. Struct. **632**, ix (2003)
4. Maseras, F., Morokuma, K.: IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. J. Comput. Chem. **16**, 1170–1179 (1995)
5. De Vita, A., Car, R.: A novel scheme for accurate MD simulations of large systems. Mater. Res. Soc. Symp. Proc **491**, 473 (1998)
6. Csányi, G., Albaret, T., Payne, M.C., De Vita, A.: "Learn on the fly": A hybrid classical and quantum-mechanical molecular dynamics simulation. Phys. Rev. Lett. **93**, 175,503 (2004)

7. Csányi, G., Albaret, T., Moras, G., Payne, M.C., De Vita, A.: Multiscale hybrid simulation methods for material systems. J. Phys. Condens. Matter. **17**, R691–R703 (2005)
8. Moras, G., Csanyi, G., Payne, M., De Vita, A.: A novel molecular dynamics approach to large semiconductor systems. Phys. B Condens. Matter. **376–377**, 936–939 (2006)
9. Csányi, G., Moras, G., Kermode, J.R., Payne, M.C., Mainwood, A., De Vita, A.: Multiscale modeling of defects in semiconductors: A novel molecular dynamics scheme. In: D.A. Drabold, S.K. Estreicher (eds.) Theory of Defects in Semiconductors, pp. 193–212. Springer, Heidelberg (2007)
10. Broughton, J.Q., Abraham, F.F., Bernstein, N., Kaxiras, E.: Concurrent coupling of length scales: Methodology and application. Phys. Rev. B **60**, 2391–2403 (1999)
11. Stillinger, F.H., Weber, T.A.: Computer simulation of local order in condensed phases of silicon. Phys. Rev. B. **31**, 5262 (1985)
12. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: Numerical Recipes: The Art of Scientific Computing, 2nd edn. Cambridge University Press, Cambridge (UK) (1992)
13. Kermode, J.R.: Multiscale hybrid simulation of brittle fracture. Doctor of Philosophy Thesis, University of Cambridge (UK) (2008)
14. Moras, G.: Modelling the growth of hydrogen-induced platelets in silicon with the "Learn on the fly" multiscale molecular dynamics technique. Doctor of Philosophy Thesis, University of London (UK) (2008)
15. Kwon, I., Biswas, R., Wang, C., Ho, K., Soukoulis, C.: Transferable tight-binding models for silicon. Phys. Rev. B. **49**, 7242 (1994)
16. Segall, M.D., Lindan, P.J.D., Probert, M.J., Pickard, C.J., Hasnip, P.J., Clark, S.J., Payne, M.C.: First-principles simulation: Ideas, illustrations and the CASTEP code. J. Phys. Condens. Matter. **14** 11, 2717–2743 (2002)
17. Oyama, N., Ohno, T.: Migration processes of the partial dislocation in silicon. Phys. Rev. Lett. **93**, 195,502 (2004)
18. Bulatov, V.V., Justo, J.F., Cai, W., Yip, S., Argon, A.S., Lenosky, T., Dekoning, T., Diaz de la Rubia, T.: Parameter-free modelling of dislocation motion: The case of silicon. Philos. Mag. A **81**, 1257–1281 (2001)
19. Baillin, X., Pelissier, J., Bacmann, J.J., Jacques, A., George, A.: Dislocation transmission through sigma=9 symmetrical tilt boundaries in silicon and germanium .1. Insitu observations by synchrotron x-ray topography and high-voltage electron-microscopy. Philos. Mag. A **55**, 143–164 (1987)
20. Jacques, A., George, A., Baillin, X., Bacmann, J.J.: Dislocation transmission through sigma=9 symmetrical tilt boundaries in silicon and germanium .2. Dynamic and crystallographic analysis. Philos. Mag. A **55**, 165–181 (1987)
21. Arias, T.A., Joannopoulos, J.D.: Ab initio theory of dislocation interactions: From close-range spontaneous annihilation to the long-range continuum limit. Phys. Rev. Lett. **73**, 680–683 (1994)
22. Huang, Y.M., Spence, J.C.H., Sankey, O.F.: Dislocation kink motion in silicon. Phys. Rev. Lett. **74**, 3392–3395 (1995)
23. Inglis, C.E.: Stresses in a plate due to the prescence of cracks and sharp corners. Trans. Inst. Naval. Archit. **55**, 219 (1913)
24. Griffith, A.A.: The phenomena of rupture and flow in solids. Philos. Trans. R. Soc. Lond. A **221**, 163 (1921)
25. Broberg, K.B.: Cracks and Fracture. Academic Press, San Diego, CA (1999)
26. Freund, L.B.: Dynamic Fracture Mechanics. Cambridge University Press, Cambridge, UK (1990)
27. Lawn, B.: Fracture of Brittle Solids, 2nd Edition. Cambridge University Press, Cambridge, UK (1993)

28. Lin, I.H., Thomson, R.: Cleavage, dislocation emission, and shielding for cracks under general loading. Acta Metall. **34**, 187–206 (1986)
29. Irwin, G.R.: Fracturing of Metals, pp. 147–166. American Society for Metals, Cleveland, OH (1948)
30. Bowler, D.R., Fearn, M., Goringe, C.M., Horsfield, A.P., Pettifor, D.G.: Hydrogen diffusion on Si(001) studied with the local density approximation and tight binding. J. Phys. Condens. Matter. **10**, 3719 (1998)
31. Kermode, J.R., Albaret, T., Sherman, D., Bernstein, N., Gumbsch, P., Payne, M.C., Csányi, G., De Vita, A.: Low-speed fracture instabilities in a brittle crystal. Nature **455** (7217), 1224–1227 (2008)
32. Li, J.: AtomEye: an efficient atomistic configuration viewer. Model. Simulat. Mater. Sci. Eng. **11**, 173 (2003)
33. Bruel, M.: Silicon on insulator material technology. Electron. Lett. **31**, 1201–1202 (1995)
34. Reboredo, F.A., Ferconi, M., Pantelides, S.T.: Theory of the nucleation, growth and structure of hydrogen-induced extended defects in silicon. Phys. Rev. Lett. **82**, 4870–4873 (1999)
35. Martsinovich, N., Suarez Martinez, I., Heggie, M.I.: First principles modelling of (100) H-induced platelets in silicon. Phys. Status Solidi C. **6**, 1771–1780 (2005)
36. Swadener, J.G., Baskes, M.I., Nastasi, M.: Stress-induced platelet formation in silicon: A molecular dynamics study. Phys. Rev. B. **72**, 201,202 (2007)
37. Terreault, B.: Hydrogen blistering of silicon: Progress in fundamental understanding. Phys. Stat. Sol. A. **204**, 2129–2184 (2007)
38. Moras, G., Colombi Ciacchi, L., Csányi, G., De Vita, A.: Modelling (100) hydrogen-induced platelets in silicon with a multi-scale molecular dynamics approach. Phys. B Condens. Matter **401–402**, 16–20 (2007)
39. Frauenheim, T., Seifert, G., Elsterner, M., Hajnal, Z., Jungnickel, G., Porezag, D., Suhai, S., Scholz, R.: A Self-consistent charge density-functional based tight-binding method for predictive materials simulations in physics, chemistry and biology. Phys. Status Solidi B **217**, 41 (2000)

CHAPTER 2

MULTISCALE MOLECULAR DYNAMICS AND THE REVERSE MAPPING PROBLEM

BERND ENSING[1] AND STEVEN O. NIELSEN[2]

[1]*Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands,e-mail: b.ensing@uva.nl*

[2]*Department of Chemistry, University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080, USA,e-mail: steven.nielsen@utdallas.edu*

Abstract: Multiscale techniques are becoming increasingly important for molecular simulation as a result of interest in increasingly complex problems involving events occurring over multiple time and length scales. Here, inspired by the success of the multiscale quantum mechanics/molecular mechanics (QM/MM) methods, we introduce a hybrid, adaptive resolution, multiscale molecular dynamics method that combines accurate, atomistic, modeling of key regions of the system with a coarse-grained modeling of the remainder of the system. Hybrid multiscale methods must solve the interfacial hand-shaking problem of coupling together different levels of description in different spatial regions of the system; in addition, to implement an adaptive resolution algorithm to correctly model diffusive systems, one must have a procedure in place to dynamically change the representation of a molecule, either from a finer to a coarser level or vice versa. We propose a solution to these problems through a detailed energy analysis and the use of a rotational dynamics to align molecular fragments. The algorithms we propose significantly advance the state-of-the-art and should serve to spur significant advances in our ability to model complex chemical systems.

Keywords: Multiscale, Adaptive resolution, Coarse-graining, Molecular dynamics, Reverse mapping, Rotational dynamics

2.1. INTRODUCTION

In numerical analysis and computer simulations, multiscale techniques are used where possible to obtain higher accuracies for lower computational cost. Weather forecast simulations and fluid dynamics calculations, for example, often make use of solving Navier-Stokes-like differential equations on discrete grids with a variable length spacing, so-called multigrids. The choice of the length scale is arbitrary; perhaps a particular part of the calculation is of greater interest and is therefore performed at a higher resolution or accuracy than the remainder. But more often, the

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 25–59.
DOI 10.1007/978-1-4020-9785-0_2,

length scales used reflect those of the underlying physics of the simulated process. In the case of weather prediction, mountainous regions with rapidly varying topography require pressure and temperature evaluations on a denser grid than do flat regions.

The properties of materials are governed by processes that take place over a vast range of length and time scales. Creep in polymers and glasses is a very slow process that can easily have relaxation constants on the order of reciprocal years. On the other hand, the atomic motions in the same materials take place on the femtosecond time scale. The electronic motions at the onset of a chemical reaction are even faster, while the chemical reaction itself might take place only a few times per second within a certain amount of material, resulting in a time scale ratio of 15 orders of magnitude. This inherent multiscale character of phenomena in materials is seen also in the length scales. A charge transfer or proton transfer chemical reaction can be a very localized process, taking place within a region of radius ten Ångstroms. The same reaction taking place in the active site of an enzyme involves the catalytic effect of the tens of nanometers-sized protein environment, while in general biological processes taking place in the cell, such as signal transduction or gene expression, are often intricately governed by long-range changes in the environment.

For the modeling of molecular phenomena with such inherent multiscale character, new developments have been made to extend existing simulation techniques that could otherwise only be applied to rather limited ranges of application. Accurate quantum mechanical ab initio methods allow for electronic structure calculations using a large supercomputer involving tens to hundreds of atoms for tens of picoseconds. Neglecting electronic structure, the behavior of molecular systems of up to a million atoms can be simulated for hundreds of nanoseconds using classical molecular dynamics techniques. Modeling of even larger systems or of processes that take place on even longer time scales requires one to relinquish an atomistic representation and simulate the motion of effective particles that each comprise several atoms; this is the domain of coarse-grained and mesoscale methods.

However, there is another choice: one can use multiscale techniques. For example in quantum chemistry, one can include the extended environment of an electronic process using embedding techniques [1, 2], continuum models [3–5] or hybrid quantum mechanics and classical forcefield (QM/MM) techniques [6]. In classical molecular dynamics simulations implicit solvent models can be employed to reduce the computation cost (with a concomitant loss of accuracy) of including for example a protein environment. And more recently, so-called hybrid multiscale methods have been developed that mix an atomistic molecular dynamics treatment of part of a system with a lower resolution treatment of the rest [7–15].

Popularity of the term *multiscale* has led the word to be used in various contexts and with an increasing (multi-) scale of meanings. Here, we will distinguish between two types of simulation techniques termed multiscale, namely (1) those in which methods of different accuracy and scale are used sequentially [16, 17], and (2) those in which methods of different accuracy and scale are used simultaneously [7–12, 18–21]. The first type makes use of a relatively accurate, high

resolution, method to parameterize the less accurate, low resolution, method which is then used for the actual simulation. In a sense, this type of multiscale method includes practically all semi-empirical and empirical methods, such as MD simulations that employ forcefields that were first parameterized using quantum chemical (e.g. Hartree-Fock) calculations. Instead, the latter type of multiscale techniques are based on treating different parts of a system with different resolutions and include for example multi-grid methods, multi-timestep techniques, certain Hamiltonian switch or replica exchange methods and hybrid methods. The multiscale method that is the subject of this book chapter is a hybrid method that allows key parts of the system to be treated at a high, atomistic, level of resolution while the rest of the system is modeled at a lower, coarse-grained, level of resolution [7].

Hybrid multiscale methods must solve the interfacial hand-shaking problem of coupling together different levels of description in different spatial regions of the system. If the shape or position of these spatial regions is changing in time, or if particles are allowed to move between the spatial regions, a special mechanism must be introduced into the multiscale method to allow particles to dynamically adapt their representation. The "Learn On The Fly" method is an example of such an adaptive hybrid method that has been successfully applied to model the propagation of a crack in a brittle solid, in which only the atoms in the advancing crack tip region are modeled at the QM tight-binding level of theory and the rest are treated with a classical forcefield representation [22]. In this case, a one-to-one mapping exists between the atoms in the quantum representation and those in the classical representation (i.e. only the number of electronic degrees of freedom differs). Bridging between an atomistic representation and a coarse-grained representation, in which each coarse-grained particle comprises several atoms, however, raises the additional difficulty that the mapping between the two representations is no longer trivial. In particular, the so-called *reverse* mapping of moving from the lower, coarse-grained, resolution to the higher, atomistic, resolution is fraught with difficulty as it requires the generation of information.

The need for adaptive boundaries within multiscale modeling methods becomes more urgent as we move from hybrid high-level QM/low-level QM methods and QM/MM methods, along the accessible time and length scales, to hybrid atomistic/coarse-grained methods. Whereas often a static boundary suffices on the relatively small, tens of picoseconds, timescale accessible to QM/MM, as shown for example by its success in modeling enzymatic chemistry, instead adaptive boundaries become crucial on the microsecond timescale domain of coarse-grained models, and therefore also hybrid atomistic/coarse-grained methods, which are particularly developed to study diffusive behavior in complex fluids, such as polymers, proteins, and amphiphilic assemblies. Adaptive boundaries are thus expected to be essential in hybrid atomistic/coarse-grained studies of a wide range of applications, such as protein folding, protein-protein interaction, permeation in (trans-membrane) ion-channels and pores, diffusion and adsorption on surfaces and open-framework materials, permeation through polymer electrolytes, self-assembly of nano-materials, and so on.

In the remainder of this book chapter, we describe the construction of a hybrid multiscale molecular dynamics method that bridges regions of high (atomistic) resolution and regions of low (coarse-grained) resolution, paying special attention to the reverse mapping problem that needs be overcome in order to make the method adaptive. But first we give a brief introduction to atomistic and coarse-grained molecular dynamics and on the forward and reverse mapping between these representations.

2.1.1. Atomistic and Coarse-Grained Molecular Dynamics

Atomistic and coarse-grained molecular dynamics simulations are particle-based methods in which conservative forces, and also sometimes dissipative and random forces, are used to evolve the particles in time; this time evolution is called a trajectory. Such classical trajectories in molecular systems can be computed on present day computers for systems containing $10^5 - 10^6$ particles for as long as about 10^8 discrete time steps. If these particles are chosen to be the atoms, the requirement to accurately sample the molecular vibrations will set the maximum total simulation time to be less than a microsecond. The most cpu-intensive part of the computation is the evaluation of the non-bonded van der Waals and electrostatic interactions, which are typically approximated by pairwise additive 2-body potentials, yielding in principle a quadratic scaling of the computational effort with system size.

For certain long-time or large length scale phenomena, the fastest molecular vibrations are irrelevant, making it desirable to average out these high frequency fluctuations a priori and model directly the representative (coarse grained) particles that move on the mesoscopic length and time scale. To model these large scale motions in complex fluids such as polymers, colloids, surfactants and bio-molecular assemblies, mesoscopic simulation methods have been developed, for instance Dissipative Particle Dynamics [23], Langevin Dynamics [24], and Brownian Dynamics [25], in which the local atomic rattling is simplified to random noise and dissipation terms. The "potential energy surface" on which these coarse-grained particles move can in principle be computed from the high-resolution representation of the system by integrating over all irrelevant fast-frequency degrees of freedom r:

$$U(R) = -k_B T \ln \int dr\, e^{-V(R,r)/k_B T} \tag{2-1}$$

with k_B Boltzmann's constant, T the absolute temperature, and V the potential that governs the fundamental interactions in the system. The resulting effective potential felt by the coarse-grained particles, $U(R)$, is actually a free energy surface that is also a function of the thermodynamic variables that define the state at which $U(R)$ is evaluated, for example the temperature and the pressure in an isothermal, isobaric ensemble. Free energy methods, such as the metadynamics method [26, 27], allow for the evaluation of a free energy landscape of a (very) small set of slow variables, which can then be used in a Langevin dynamics to explore the kinetics. However, for the construction of a (many-) particle based coarse-grained model, calculation

of the exact potential $U(R)$ is not feasible in practise and people have devised other approximate methods to build coarse-grained potentials, or forcefields.

The level of coarse-graining applied here in the context of hybrid multiscale molecular dynamics is in between atomistic and mesoscopic. At this level, coarse-grained particles represent approximately the chemical functional groups of molecules, containing in the order of ten atoms each [28]. That way, the same machinery as used for atomistic molecular dynamics can be used for the coarse-grained molecular dynamics, including for example harmonic bond and bend functions and non-bonded van der Waals-like and electrostatic interactions that make up the forcefield. Critically, this lets us use existing atomistic molecular dynamics software to carry out the coarse grained simulations, so that we do not have to worry about writing efficient, parallelized simulation code. Moreover, at this level the molecular shape is preserved which conceptually allows for a mapping between the atomistic and coarse-grained representations of the system.

Motivated by Henderson's theorem [29], which states that there exists a one-to-one mapping between measured pair-correlation functions and the underlying potential energy function in the case that the latter is a sum of pair-interactions, strategies to construct coarse-grained forcefield are often based on constructing effective pair-potentials from pair-correlation functions and potentials of mean force obtained from atomistic simulations [30–34]. These approaches work well in cases where 3-body and higher correlations are weak. The resulting effective potentials that reproduce the target distributions are not unique however, which leaves room to simultaneously match other target properties, for example experimental densities, surface tensions, heats of vaporization, and so forth. Jain and co-workers showed that such inclusion of thermodynamic target properties in the optimization procedure has the added advantage of increased convergence [33]. A different method to build coarse-grained potentials is by trying to match the effective forces on the coarse-grained degrees of freedom within an atomistic simulation [35–37].

It is beyond the scope of the current multiscale topic to discuss the art of coarse-graining in more depth. It is however important, and hopefully obviously so, to ensure that the low-resolution and high-resolution models of the system, merged in an hybrid multiscale method, represent the same thermodynamic state point. The limited transferability of coarse-grained forcefields (as mentioned for $U(R)$ in Eq. 2-1) as compared to atomistic forcefields, might therefore require one to re-optimize the coarse-grained potentials for new simulation conditions (e.g. a different temperature or pressure) to ensure the same chemical potential in the different low and high resolution regions and avoid spurious density fluctuations.

2.1.2. Mapping Between Different Representations, or the Reverse Mapping Problem

In order to implement an adaptive resolution algorithm, one must have a procedure in place to dynamically change the representation of a molecule, either from a finer to a coarser level or vice versa. The "forward" direction, namely where one coarsens the representation of a molecule, is straightforward because one merely

throws away information; for example by replacing the atomic coordinates of a collection of atoms with the coordinates of their center of mass (COM). On the other hand, the "reverse" mapping, in which one resolves a molecule into finer detail, is problematic because it requires the creation of information. Indeed, in a recent review article on multiscale modeling, de Pablo and Curtin say that "a persistent challenge remains that of reverse mapping, that is, of restoring some of the details after they have been blurred away through an averaging procedure" [38].

Here we propose a solution to the reverse mapping problem. We associate a frozen atomistic fragment with each coarse-grained site, and rigidly rotate these fragments about their COM in accordance with an energy function designed to maintain a low-energy atomistic-level representation of the system. In this manner the coarse-grained system has, superimposed on it, a globally unfrustrated atomistic configuration which is prepared for reverse mapping. Indeed, only a local relaxation needs to be further performed on the degrees of freedom which were frozen (the degrees of freedom within each fragment). This is accomplished, for instance, by running a short atomistic simulation or by using a healing region as will be described below. This method is efficient because the atomistic fragments are treated as rigid bodies with no internal degrees of freedom. This means that there are only three degrees of freedom per coarse-grain site to evolve, namely an element of the special orthogonal Lie group SO(3). The mathematical details of this treatment are given below.

2.2. ADAPTIVE MULTISCALE MOLECULAR DYNAMICS

Due to the reverse mapping problem described above, constructing a meaningful adaptive multiscale molecular dynamics algorithm is not trivial. First of all, the number of degrees of freedom continuously changes during such a simulation as low-resolution (coarse-grained) particles break up into their high-resolution components (atoms) when they enter the high-resolution region and vice versa when they leave. During these transitions the number of pair-interactions also changes, so that both the total potential energy and the total kinetic energy show spurious fluctuations with the changing numbers of particles in the low-resolution and high-resolution regions. What thermodynamic ensemble would such an adaptive multiscale simulation sample? Secondly, and this is precisely the reverse mapping problem, a coarse-grained particle entering the high-resolution region has to be replaced by its fine-grained counterpart, which requires the generation of information: how can one generate suitable equilibrium positions and velocities for these atoms in harmony with the other atoms in the high-resolution region?

In this section, a possible solution is given to the above difficulties that come with the construction of an algorithm for adaptive multiscale molecular dynamics, or hybrid MD. The algorithm that we discuss is aimed to be simple, robust, and have certain desired properties such as conservation of linear momentum and conservation of total energy. This algorithm is built in two stages. Starting from an atomistic (i.e. the high resolution) representation of the entire system and a

(forward) mapping of groups of atoms into coarse-grained beads, the atomistic pair-interactions that span atoms belonging to different coarse grained beads are replaced by coarse-grained pair-interactions in the coarse-grained region. The details of such a coarse-graining of interactions is presented in the following subsection where we discuss the treatment of the coupling between atoms in the atomistic and coarse-grained regions. The second stage then involves "freezing" of the atoms in the coarse-grained region inside their coarse-grained beads so that the dynamics of the atoms can be replaced by a coarse-grained dynamics. This is presented in the subsequent subsection. Next, we present an illustrative case study of a hybrid molecular dynamics (hybrid MD) simulation of liquid methane. We end with a brief discussion of variations of and alternatives to this multiscale algorithm that have recently appeared in the literature.

2.2.1. Stage 1: Coupling Atomistic and Coarse-Grained Regions

Let us consider a molecular system that we wish to separate into two regions, one of which contains the part that we want to treat in atomistic detail and the other which contains the rest to be treated at a lower, coarse-grained, resolution. This section describes the first of the two stages in which such a multiscale treatment can be constructed, leading to an adaptive algorithm that allows particles to move between the two regions and adapt their representation on the fly.

But first, let us recall that a conventional molecular dynamics treatment of the entire system at the atomistic level maintains the micro-canonical *NVE* ensemble as governed by the conservation laws of Hamilton's equations. In this ensemble the number of atoms, N, the volume, V, and the total energy, E, remain constant. The total energy is taken to be the sum of the total kinetic energy, T, and the total potential energy, V, summed over all atoms, i:

$$E = T + V = \sum_{i}^{N} \frac{1}{2} m_i v_i^2 + V(r^N) \tag{2-2}$$

in which m_i and v_i are the mass and velocity of atom i respectively, and the potential depends on all the positions r_i.

Other ensembles can be generated, for example the canonical *NVT* ensemble or the isobaric-isothermal *NPT* ensemble by invoking a thermostat or a barostat plus a thermostat. Note that now the total energy is no longer constant. However, also in these ensembles a conserved quantity such as the total energy can be recovered by cleverly bookkeeping the energy flows between the molecular system and the external variables introduced to control the temperature and/or the pressure. The Nosé-Hoover (chain) thermostat [39, 40], the Parrinello-Rahman barostat [41], and the recently introduced stochastic velocity rescaling thermostat by Bussi [42] are all good examples of external control mechanisms that provide the desired ensemble and recover a conserved total energy even though the dynamics is not strictly Hamiltonian.

Conserved quantities play a very important role in molecular simulations. Monitoring the conservation of total energy yields the first and foremost assessment of the quality of the integration of the equations of motion during the simulation. Whether the computer code contains a bug, or the simulation time step is too large, or the initial system configuration contains overlaps between atoms, the problem is always first seen in the (non-) conservation of the total energy. Given this importance, it seems like a good idea to focus on energy conservation while constructing a multiscale algorithm.

The construction of a multiscale algorithm begins from the atomistic representation of the entire system and some mapping in which groups of atoms are lumped into coarse-grained beads, which can for example be positioned at the centers of mass of the atomic groups they represent. The first stage of coarse-graining the low-resolution part of the system consists of replacing all atomistic interactions that span different beads by coarse-grained interactions. The second stage consists of replacing the atoms in this region by their coarse-grained bead representation, which is dealt with in the next subsection. Specifically, for now, the atomistic bonded (and non-bonded) interactions between atoms i and j that belong to the same coarse-grained bead α are maintained, and only the interactions between atoms i and j that belong to different beads α and β are replaced by coarse-grained potentials. The energy function of such a system reads:

$$E = \sum_{i}^{n+m} \frac{1}{2} m_i v_i^2 \; + \; V^{\mathrm{A}}(r^n) + V^{\mathrm{CG}}(R^M) + V^{\mathrm{mix}}(R^N, R^M) \tag{2-3}$$

$$\begin{aligned} V^{\mathrm{A}}(r^n) &= \sum_{i=1}^{n-1} \sum_{j>i}^{n} \Phi^{\mathrm{A}}(r_i, r_j) \\ V^{\mathrm{CG}}(R^M) &= \sum_{\alpha=1}^{M-1} \sum_{\beta>\alpha}^{M} \Phi^{\mathrm{CG}}(R_\alpha, R_\beta) \\ V^{\mathrm{mix}}(R^N, R^M) &= \sum_{\alpha=1}^{N} \sum_{\beta=1}^{M} \Phi^{\mathrm{CG}}(R_\alpha, R_\beta) \end{aligned}$$

so that the n atoms in the atomistic region interact with each other through an atomistic potential or forcefield $V^{\mathrm{A}}(r^n)$ while the remaining m atoms in the coarse-grained region interact with each other through a coarse-grained potential or forcefield V^{CG} working on the M centers of mass, R, of the atomic fragments corresponding to beads. Here the potential terms are expressed as sums of pair-potentials, Φ. The effective force on such a center of mass is then distributed, mass weighted, over the atoms that belong to the coarse-grained bead. The coupling between atoms in different regions is governed by the last term $V^{\mathrm{mix}}(R^N, R^M)$, which is also the coarse-grained potential working on the centers of mass and therefore not different from V^{CG}, except that then α runs over the N beads in the atomistic region, instead of the $M - 1$ beads in the coarse-grained region.

The total energy in Eq. (2-3) is conserved, however this multiscale setup is not yet adaptive so that after some simulation time, due to diffusion, the atoms "feeling" fully atomistic interactions will start to mix with those atoms from the coarse-grained region that only feel their environment through coarse-grained interactions, which is not what we set out to do. To make the method adaptive and have the atoms switch their interaction when they cross user-defined regions, we introduce a spatial order parameter q to distinguish the two regions and we denote the boundary between the two regions by $q^{\dagger}$. With this order parameter all atoms can be attributed a label, s, which we shall call the (amount of) "coarse-grained character", and which equals zero for atoms in the (atomistic) region of $q < q^{\dagger}$ and one for atoms in the (coarse-grained) region with $q > q^{\dagger}$, as shown in the upper panel (panel A) in Figure 2-1.

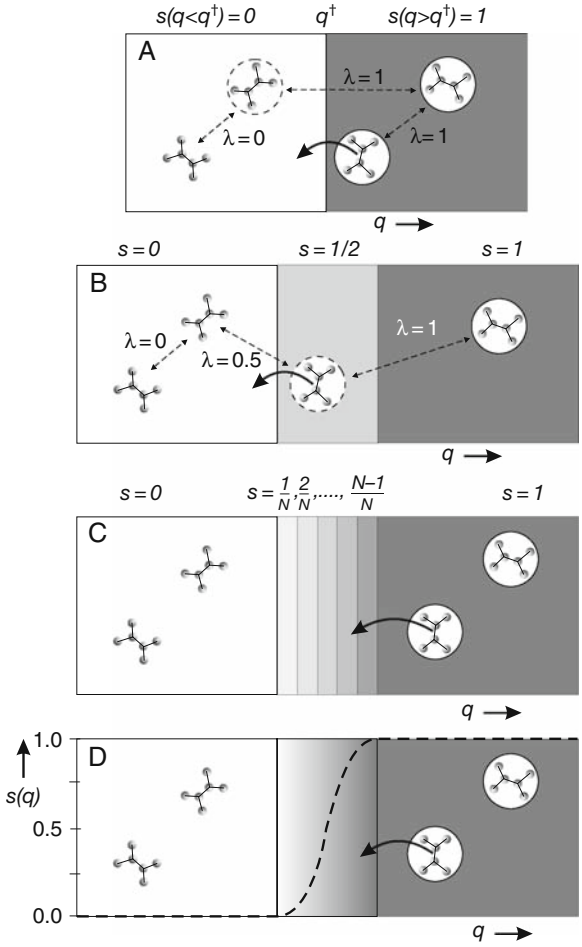


Figure 2-1. An intermediate "healing region" that smooths the coupling between the high-resolution and low-resolution regions is introduced in three simple steps

Next, the amount of coarse-grained character, s, is employed to make the potential function adaptive:

$$V^{\mathrm{A}} + V^{\mathrm{CG}} + V^{\mathrm{mix}} = \sum_{\alpha\beta} \left(\lambda_{\alpha\beta} \Phi^{\mathrm{CG}}_{\alpha\beta} + (1 - \lambda_{\alpha\beta}) \sum_{\substack{i \in \alpha \\ j \in \beta}} \Phi^{\mathrm{A}}_{ij} \right) + \sum_{\alpha} \sum_{i,j \in \alpha} \Phi^{\mathrm{A}}_{ij} \qquad (2\text{-}4)$$

$$\lambda_{\alpha\beta} = \max(s_\alpha, s_\beta)$$

Here, the previous potential energy terms are replaced by the two terms on the right. The first term accounts for all interactions between different coarse-grained beads α and β, which are either taken as the coarse-grained pair-potential Φ^{CG} when $\lambda_{\alpha\beta} = 1$ or as the sum of atomistic pair-potentials Φ^{A} between atoms i belonging to bead α and atoms j belonging to bead β when $\lambda_{\alpha\beta} = 0$. The value of the scaling function, $\lambda_{\alpha\beta}$, being zero or one, is determined by the maximum coarse-grained character s between the two interacting beads α and β, as illustrated in Figure (2-1A). Using this scaling factor, the same multiscale behavior is obtained as before, namely, atoms in the atomistic region interact through the atomistic forcefield while they feel the particles in the coarse-grained region through the coarse-grained forcefield, and the atoms in the coarse-grained region "see" all other atoms, in both regions, as coarse-grained particles. The difference is now that this interaction automatically adapts when atoms move from one region to the other. The last term in Eq. (2-4) accounts for all bonded (and non-bonded) interactions within each bead. Note also that this equation is trivially generalized for interactions other than pair-interactions, such as bending (or dihedral) potentials, by taking for lambda the maximum s-value of the now three (or four) interacting atoms.

By making the multiscale approach adaptive, we have introduced two new problems: one, the total energy is no longer conserved, and, two, also the forces are discontinuous when particles cross the boundary at $q^{\dagger}$. In other words, as long as the particles stay in their initial regions the energy is conserved and the dynamics is well-behaved, but every time a particle crosses $q^{\dagger}$ the total energy and the forces will display a jump. The jump in the energy when the atoms of bead α cross from the atomistic region to the coarse-grained region equals the difference between the coarse-grained potential and the atomistic potential due to all other atoms:

$$\Delta U^{\mathrm{A/CG}}_{\alpha} = \sum_{\beta} \Delta\lambda_{\alpha\beta} \left(\tilde{\Phi}^{\mathrm{CG}}_{\alpha\beta} - \sum_{\substack{i \in \alpha \\ j \in \beta}} \tilde{\Phi}^{\mathrm{A}}_{ij} \right) \qquad (2\text{-}5)$$

where $\Delta\lambda_{\alpha\beta}$ is the change of λ for each pair-interaction, which equals zero for all β particles in the coarse-grained region (as these λ remain unity), and for all other β equals either +1 if α moves from the atomistic region to the coarse-grained region, or -1 if α moves from the coarse-grained region to the atomistic region. The tilde indicates the value of the interaction, Φ, at the moment of boundary crossing. By simply adding this term, $\Delta U_\alpha^{\mathrm{A/CG}}$, every time a particle crosses between regions, the total energy is recovered as a conserved quantity. Fixing the spurious jumps in the forces, however, will require more than just this simple bookkeeping. This is what we set out to do next.

Some improvement can be envisioned by introducing an intermediate "healing region" between the atomistic region and the coarse-grained region, in which particles are attributed a mixed atomistic/coarse-grained character of $s = 0.5$. This is illustrated in Figure 2-1, panel B. Particles in this healing region interact with other "hybrid" particles in this region through a potential that is the sum of the atomistic interactions and the coarse-grained interactions, both scaled by $\lambda = 0.5$, according to Eq. (2-4). The same type of interaction is felt between these hybrid particles in the healing region and particles in the atomistic region (as $\max(s_1, s_2) = 0.5$), while the interaction between the hybrid particles and particles in the coarse-grained region is purely the coarse-grained Φ^{CG} (i.e. $\max(s_1, s_2) = 1$).

Similar to the previous dual-scale setup, the evolution of the atomistic dynamics can be performed on this, now three-region, system, where the potentials are switched accordingly every time a particle crosses the boundary between the atomistic region and the healing region or the boundary between the coarse-grained region and the healing region. In other words, the bookkeeping $\Delta U_\alpha^{\mathrm{A/CG}}$ now supplements the energy function when a particle α switches between $s = 0$ and $s = 0.5$ and when a particle switches between $s = 0.5$ and $s = 1$, to correct the jumps in the total energy. These corrections are smaller than with the previous single region boundary because $\Delta\lambda$ (in Eq. (2-4)) now only amounts to $+\frac{1}{2}$ or $-\frac{1}{2}$, and with that also the jumps in the forces are somewhat smaller.

Further improvement can thus be made by dividing the healing region into several, say $N-1$, intermediate sub-regions, as shown in panel C of Figure 2-1. Particles in each sub-region k could then be attributed a stepwise increasing coarse-grained character of $s = k/N$, starting from $k = 0$ in the fully atomistic region to $k = N$ in the fully coarse-grained region. Again, a particle in sub-region k interacts with all particles that find themselves in the same sub-region k or in regions smaller than k with a hybrid potential, given by Eq. (2-4), with scaling factor $\lambda = k/N$, whereas this particle interacts with all other particles in higher regions with a hybrid potential with a larger scaling factor that is determined by the other particle. For a fixed total healing region width with larger and larger numbers of intermediate sub-regions, N, the intermediate regions become narrower in spatial extent and particles are more frequently found to cross one or another region boundary. A particle α crossing from region k to region $k + 1$ requires an update of the bookkeeping term $\Delta U_\alpha^{\mathrm{A/CG}}$ equal to:

$$\Delta U_\alpha^{\mathrm{A/CG}} = \sum_\beta \left(\frac{k+1}{N}\Phi_{\alpha\beta}^{\mathrm{CG}} - \frac{k}{N}\Phi_{\alpha\beta}^{\mathrm{CG}} - \frac{N-k-1}{N}\sum_{\substack{i\in\alpha\\ j\in\beta}}\Phi_{ij}^{\mathrm{A}} + \frac{N-k}{N}\sum_{\substack{i\in\alpha\\ j\in\beta}}\Phi_{ij}^{\mathrm{A}} \right)$$
$$= \sum_\beta \left(\frac{1}{N}\Phi_{\alpha\beta}^{\mathrm{CG}} - \frac{1}{N}\sum_{\substack{i\in\alpha\\ j\in\beta}}\Phi_{ij}^{\mathrm{A}} \right) \tag{2-6}$$

which is not different from the bookkeeping term defined by Eq. (2-5) with $\Delta\lambda = \pm 1/N$. In the case of very narrow sub-regions, a particle moving in any direction that is not strictly perpendicular to q will cross several sub-region boundaries, which requires summing over as many bookkeeping terms $\Delta U_\alpha^{\mathrm{A/CG}}$.

Now, taking the limit of the number of sub-regions going to infinity,

$$\Delta U_\alpha^{\mathrm{A/CG}} = \lim_{N\to\infty} \sum_{k_q}^{k_{q'}} \sum_\beta \Delta\lambda_{\alpha\beta} \left(\Phi_{\alpha\beta}^{\mathrm{CG}} - \sum_{\substack{i\in\alpha\\ j\in\beta}}\Phi_{ij}^{\mathrm{A}} \right) \tag{2-7}$$

the sum over boundaries for a particle moving from sub-region k at q to sub-region k' at q' can be replaced by an integral:

$$\Delta U_\alpha^{\mathrm{A/CG}} = \int_q^{q'} dq \sum_\beta \frac{d\lambda_{\alpha\beta}}{dq} \left(\Phi_{\alpha\beta}^{\mathrm{CG}} - \sum_{\substack{i\in\alpha\\ j\in\beta}}\Phi_{ij}^{\mathrm{A}} \right). \tag{2-8}$$

Of course, the linear change of s and therefore of the scaling factor, λ, of $1/N$ per intermediate region is not a requirement, and other smoothly varying functions for s can be used, as long as it switches from zero to one over the healing region and is differentiable. During a molecular dynamics simulation, it is practical to perform the integration over the full spatial trajectory rather than its projection onto q, as both the interaction potentials, Φ, and the order parameter, q, are functions of the position, r.

$$\Delta U_\alpha^{\mathrm{A/CG}} = \int dr \sum_\beta \frac{d\lambda_{\alpha\beta}}{dq}\frac{\partial q}{\partial r} \left(\Phi_{\alpha\beta}^{\mathrm{CG}} - \sum_{\substack{i\in\alpha\\ j\in\beta}}\Phi_{ij}^{\mathrm{A}} \right). \tag{2-9}$$

The resulting energy function of this adaptive multiscale method is obtained from Eq. (2-3), (2-4), and the total bookkeeping energy, $\Delta U^{\mathrm{A/CG}}$, of all particles moving in the healing region. Integration of the equations of motion derived from this energy function, with a sufficiently small time step, will in principle maintain the total energy as a conserved quantity. In reference [7] an illustrative numerical calculation

is discussed of a simple model system containing a diatomic molecule which moves through a healing region. Despite the deliberately mismatched potential functions between the atomistic and coarse-grained representations, total energy conservation is recovered to remarkably high accuracy using the numerical integration of Eq. (2-9).

2.2.2. Equations of Motion

The Newtonian equations of motion of the adaptive multiscale molecular dynamics are constructed in the usual way

$$m_i \frac{\mathrm{d}^2 r_i}{\mathrm{d}t^2} = f_i = -\frac{\partial V}{\partial r_i} \tag{2-10}$$

with the force f on particle i derived from the potential,

$$V = \sum_{\alpha\beta} \left(\lambda_{\alpha\beta} \Phi^{\mathrm{CG}}_{\alpha\beta} + (1 - \lambda_{\alpha\beta}) \sum_{\substack{i\in\alpha \\ j\in\beta}} \Phi^{\mathrm{A}}_{ij} \right) + \sum_{\alpha} \sum_{i,j\in\alpha} \Phi^{\mathrm{A}}_{ij} + \Delta U^{\mathrm{A/CG}} \tag{2-11}$$

This is the adaptive multiscale, or hybrid MD, potential, which was derived in the previous section. Summarizing, the first term sums the scaled interactions between all pairs of coarse-grained particles α and β, with the scaling factor $\lambda_{\alpha\beta}$ a number between zero and one depending on the resolution (being atomistic, coarse-grained, or something in between) of the interacting particles. The second term accounts for all atomistic interactions within the coarse-grained beads. These interactions are not scaled by λ, as they are obviously not replaced by a coarse-grained interaction, and which would otherwise thus lead to disintegration of the molecules when they leave the atomistic region. The third and last term holds the bookkeeping energies (Eq. 2-9) of all particles that change their resolution in the healing region.

Derivation of the forces from the hybrid potential requires special attention to the first and last terms which contain the position dependent switching function. The force on atom i is

$$\begin{aligned} f_{i\in\alpha} &= -\frac{\partial V}{\partial r_i} \\ &= -\sum_{\beta} \left(\left\{ \lambda_{\alpha\beta} \frac{\partial \Phi^{\mathrm{CG}}_{\alpha\beta}}{\partial r_i} + \frac{\partial \lambda_{\alpha\beta}}{\partial r_i} \Phi^{\mathrm{CG}}_{\alpha\beta} \right\} \right. \\ &\quad \left. + \sum_{j\in\beta} \left\{ (1 - \lambda_{\alpha\beta}) \frac{\partial \Phi^{\mathrm{A}}_{ij}}{\partial r_i} - \frac{\partial \lambda_{\alpha\beta}}{\partial r_i} \Phi^{\mathrm{A}}_{ij} \right\} \right) \\ &\quad - \sum_{j\in\alpha} \frac{\partial \Phi^{\mathrm{A}}_{ij}}{\partial r_i} - \frac{\partial \Delta U^{\mathrm{A/CG}}}{\partial r_i} \end{aligned} \tag{2-12}$$

in which the terms in curly brackets are obtained by applying the chain rule to the scaled coarse-grained and atomistic interactions respectively, shown in Eq. (2-11) as the term in parenthesis. The first term in both of the pieces between the curly brackets we recognize as the usual interaction force, but now multiplied by the scaling factor λ and $(1-\lambda)$ respectively. The second terms resulting from the chain rule contain the derivative of the scaling factor with respect to the particle position. This derivative is non-zero in the healing region where λ changes in the direction of the order parameter q (see also panel D in Figure 2-1). This seems strange as apparently these two terms will cause a force, and thus a flux of particles, between the atomistic and coarse-grained regions, unless $\Phi^{\mathrm{CG}}_{\alpha\beta} = \sum_{ij} \Phi^{\mathrm{A}}_{ij}$.

However, the last term, the derivative of $\Delta U^{\mathrm{A/CG}}$ is just the integrand of Eq. (2-9)

$$\frac{\partial \Delta U^{\mathrm{A/CG}}}{\partial r_i} = \sum_{\beta} \left(\frac{\partial \lambda_{\alpha\beta}}{\partial r_i} \Phi^{\mathrm{CG}}_{\alpha\beta} - \sum_{j \in \beta} \frac{\partial \lambda_{\alpha\beta}}{\partial r_i} \Phi^{\mathrm{A}}_{ij} \right) \tag{2-13}$$

which cancels exactly these two spurious terms leaving only the scaled interactions plus the intra-bead interaction in the force expression:

$$f_{i \in \alpha} = -\sum_{\beta} \left(\lambda_{\alpha\beta} \frac{\partial \Phi^{\mathrm{CG}}_{\alpha\beta}}{\partial r_i} + \sum_{j \in \beta} (1 - \lambda_{\alpha\beta}) \frac{\partial \Phi^{\mathrm{A}}_{ij}}{\partial r_i} \right) - \sum_{j \in \alpha} \frac{\partial \Phi^{\mathrm{A}}_{ij}}{\partial r_i} \tag{2-14}$$

This equation contains the usual symmetry with respect to interacting particles i and j which ensures obedience to Newton's third law ($f_i = -f_j$) and conservation of momentum in the system.

2.2.3. Stage 2: Freezing the Intra-Bead Motions

In the previous two sections an adaptive multiscale dynamics approach was constructed by coupling an atomistic region, in which atoms interact through an atomistic forcefield, with a coarse-grained region in which atoms interact through a coarse-grained forcefield. Technically, however, both regions still maintain atomistic dynamics rather than a coarse-grained dynamics in the coarse-grained region. That is, the integrator propagates the atomic positions and velocities also in the coarse-grained region instead of propagating positions and velocities of the coarse-grained particles. Here, in the second stage of the multiscale method development, the atoms in the low-resolution region are replaced by coarse-grained beads.

Basically, the atoms grouped into a coarse-grained bead can be "frozen" with respect to their center of mass and replaced by the coarse-grained particle as soon as they enter the coarse-grained region from the healing region, since, from there on, the motion of the center of mass of each atomic fragment is solely governed by the coarse-grained interactions making the "internal" atomic motions irrelevant to the molecular dynamics. Such one-step-freezing is exactly what we will do here,

although alternative schemes are possible, involving for example a gradual freezing of the atomic degrees of freedom in the healing region (see Section 2.2.4 for more details).

The advantages of replacing the evolution of the atomic degrees of freedom by coarse-grained molecular dynamics are mainly economic; computation of atomic intra-bead interactions is avoided, a larger time step is allowed without violating energy-conservation, and memory storage can be saved in the coarse-grained region, which is typically the larger region. The only disadvantage of discarding the atomistic details is that of the reverse mapping problem discussed earlier: once the atomic inter-bead interactions are replaced by the coarse-grained forcefield the atomic fragments are free to rotate around their centers of mass, randomizing the atomistic details with respect to their environment outside the bead. This is why we in fact control this rotation as discussed in Section 2.3, so that we can recover the atomic details to some approximation.

The fact that, in the low-resolution region, the atoms no longer feel their environment and adopt random orientations (ignoring for the moment the possibility of SO(3) rotational dynamics) means that such groups, upon moving into the healing region and toward the atomistic region, practically always do so starting relatively high up on the atomic potential energy surface. That is, while the atomic interactions are gradually switched on while moving toward the atomistic region, the atomic degrees of freedom (have to) re-equilibrate with respect to their atomistic environment. This equilibration process of transforming from high potential energy random orientations to equilibrium energy structures generates thermal motion or heat. Note that the reverse process of switching off the atomistic interactions and allowing atomic fragments to take random orientations as molecules move toward the coarse-grained region does not require the absorption of heat. This asymmetry in heat transport with respect to movement toward the high-resolution region versus movement in the other direction means that heat is produced continuously in adaptive multiscale molecular dynamics. This heat has to be removed by a thermostat.

When particles cross the boundary between the coarse-grained region and the healing region, the atoms are replaced by a single coarse-grained particle or vice versa depending on the crossing direction. In the coarse-grained region, the atomic positions and velocities can simply be stored relative to their center of mass and thus be recovered when the particle leaves the coarse-grained region. Alternatively, atomic positions and velocities can be re-generated, for example by inserting a relaxed structure with random (Boltzmann) velocities or by copying positions and velocities from a molecule in the atomistic region. The instantaneous switching between atoms and coarse-grained particles at this boundary introduces jumps in the total energy that require two extra bookkeeping terms, namely

$$\Delta U^{\text{intraCG}} = \sum_{\alpha} \Theta\left(s_\alpha - 1\right) \sum_{i,j \in \alpha} \tilde{\Phi}_{ij}^{\text{A}} \tag{2-15}$$

$$\Delta T^{\text{A/CG}} = \sum_{\alpha} \Theta\left(s_\alpha - 1\right) \frac{1}{2} \left(m_\alpha \tilde{v}_\alpha^2 - \sum_{i \in \alpha} m_i \tilde{v}_i^2 \right) \tag{2-16}$$

The first extra bookkeeping term accounts for the atomistic intra-coarse-grained bead interactions which are no longer computed when the atoms become "frozen". Here, the tilde indicates again the values of Φ and v at the boundary crossing moment and the Heaviside step function, Θ, equals one for particles in the coarse-grained region ($s = 1$) and zero otherwise. The second equation accounts for the change in kinetic energy as the number of degrees of freedom adapts. The sum of the two terms can be seen as the instantaneous internal energy of the atomic degrees of freedom inside the coarse-grained bead, which is integrated out upon coarse-graining and thus becomes a constant in the coarse-grained region, whereupon it will be released again when switching back to the atomistic representation.

The definition of the regions (high-resolution, healing, and low-resolution) is arbitrary and can for example be chosen to be a spherical atomistic region centered on a particle of particular interest, so that the high-resolution part of the simulation follows this particle. The healing region should then also be a spherical region with the same center but with a larger radius, leaving everything outside this sphere as the low-resolution region. Alternatively, the regions can be fixed in space. In either case, after every (smallest) time step the amount of coarse-grained character, s, needs to be computed by calculating the distance between each particle and the center of the spherical regions. The coarse-grained character is used for the scaling factor in the force calculations involving all particles in the healing region. A computational saving can be made by computing s only for a list of particles that find themselves in this healing region or just outside of it (skin-regions). By taking the width of the skin-regions on either side of the healing region equal to the skin length applied in the usual neighbor list for the non-bonded interactions, the small list of particles for which s is computed needs to be rebuild only as often as the neighbor list is updated.

Summarizing, we have constructed an adaptive multicale molecular dynamics by first introducing an intermediate healing region and subsequently replacing the groups of atoms in the low-resolution region by coarse-grained beads. The total energy is recovered as a conserved quantity by adding the proper auxiliary bookkeeping terms to the energy function. Next, we will examine the behavior of this hybrid MD approach with an illustrative example.

2.2.4. Case Study 1: Liquid Methane

As an illustrative application of the adaptive multiscale dynamics method, we have performed a simulation of liquid methane at $T = 111.5$ K and atmospheric pressure. In the high resolution region, methane is treated as a fully flexible atomistic CH_4 molecule (using the CHARMM forcefield [43]) and in the low resolution region each methane molecule is modeled as a single bead using Jorgensen's united atom model [44]. A similar simulation of 8000 methane molecules has been published in reference [7], in which case the atomistic region was defined by a sphere with a radius of $R_A = 8$ Å fixed in space. Different surrounding healing regions were tried with widths varying between $R_{HR} = 1$ and 4 Å, and, not surprisingly, the largest, 4 Å wide, healing region resulted in the best performance as seen from the conservation of energy.

Here, we take instead a rectangular box with edges $L = 38.0, 38.0, 75.5$ Å, containing 1900 methane molecules subdivided into rectangular regions of different resolution. That is, the atomistic region is a slab, flanked by healing regions on both sides with a total thickness of $d_{\mathrm{A}} + 2 \times d_{\mathrm{HR}} = 44$ Å, which leaves the rest of the system as a coarse grained slab of width $d_{\mathrm{CGR}} = 31.5$ Å. See Figure 2-2 for an illustration of the system. We will compare six hybrid MD simulations in which we again vary the width of the healing regions from $d_{\mathrm{HR}} = 1$ to 6 Å (thus leaving an atomistic region of varying width between $d_{\mathrm{A}} = 42$ and 32 Å). The average number of methane molecules in the coarse-grained region was close to 790 with a standard deviation of about 17. The number of molecules in the atomistic region varied from 1063 in the $R_{\mathrm{HR}} = 1$ Å simulation to 812 in the $R_{\mathrm{HR}} = 6$ Å simulation, leaving 50–294 molecules respectively in the healing regions (see also top-right panel in Figure 2-3).

The interesting observables that illustrate the behavior of the hybrid MD method are plotted in the other three panels of Figure 2-3. Starting at the top-left panel, we see the total energy (shifted for comparison) is very well conserved for the $R_{\mathrm{HR}} = 6$ Å healing region, and showing an unstable drift in the hybrid MD simulations with healing region smaller than $R_{\mathrm{HR}} = 3$ Å. The graph at the bottom left shows the $\Delta U^{\mathrm{A/CG}}$ bookkeeping term (Eq. 2-9), which in the simulations is computed on the fly by multiplying the integrand (Eq. 2-13) by the displacement dq of the interacting particle in the healing region. Molecules moving from the coarse-grained region to the atomistic region fall quickly down from high potential energy configurations as they equilibrate into their atomistic environment, whereas molecules moving in the other direction are not *pushed up* to such high potential energy configurations. This asymmetry with respect to the direction that particles

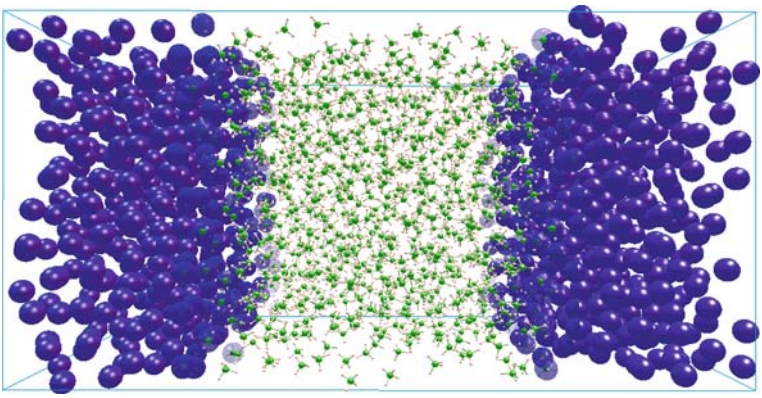


Figure 2-2. Hybrid MD snapshot of the periodic unit cell of 1900 methane molecules. In the center is the atomistic region flanked by the two halves of the coarse-grained slab, in which each CH_4 molecule is treated as a single blob (pictured in *blue*). These two regions are coupled through 6 Å wide healing regions in which the molecules smoothly change their resolution, here indicated by the varying transparency of the molecules

move (i.e. positive or negative *dq* in Eq. 2-9) means that on average the fluctuations in $\Delta U^{\mathrm{A/CG}}$ do not cancel, as reflected by the negative slopes in the plot.

The atomistic equilibration process in the healing region, driven by a gradual switching on of the atomistic interactions, forces the atoms to move and reorient. This increase in kinetic energy is counteracted by a Nosé-Hoover thermostat coupled to each particle. The bottom-right panel in Figure 2-3 shows the flow of energy from the system to the thermostat, which is larger for small healing region widths.

This case study of a hybrid MD simulation of liquid methane illustrates that the adaptive multiscale algorithm is a robust and very promising method. It also shows the importance of recovering the total energy as a conserved quantity in assessing the stability and accuracy of the simulation. Without this total energy observable, one is left with guessing, based on secondary information such as density fluctuations or other measured properties, whether the choices made for the healing region size and the time step were adequate.

2.2.5. Other Adaptive Multiscale Implementations

At this stage, we briefly outline two other approaches to (particle-based) adaptive multiscale dynamics methods, paying special attention to the differences and similarities to the hybrid MD method constructed in the previous sections.

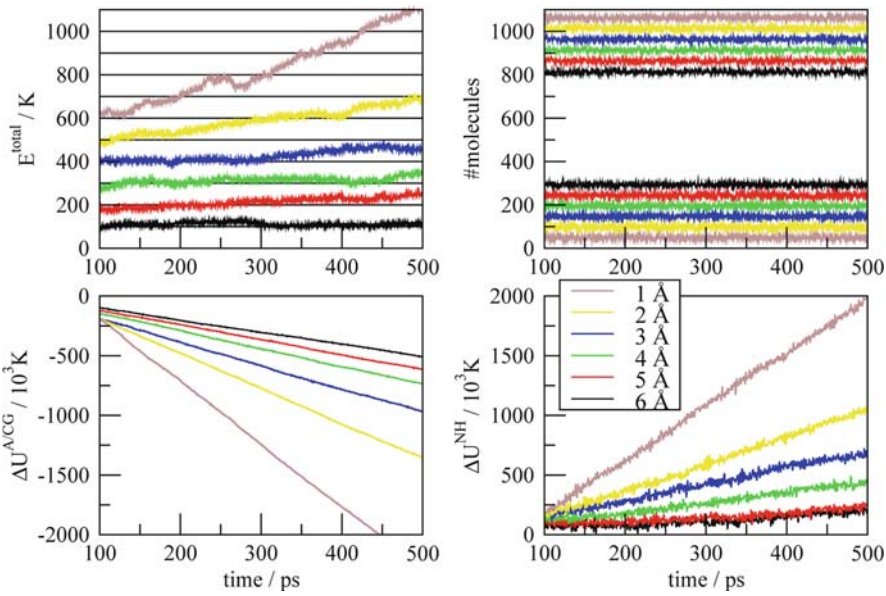


Figure 2-3. Hybrid MD simulation data for liquid methane with varying healing region widths. *Top left*: the conservation of the total energy is excellent for the 6 Å healing region and becomes unacceptable for healing regions smaller than 3 Å. *Top right*: Continuously fluctuating numbers of molecules in the atomistic region (*upper part* of the graph) and in the healing region (*lower part*). *Bottom left*: bookkeeping term $\Delta U^{\mathrm{A/CG}}$ from Eq. 2-9. *Bottom right*: continuous energy flow into the Nosé-Hoover thermostats

Praprotnik et al. [8] were the first to present an adaptive resolution dynamics scheme, only preceded by the similar adaptive Monte-Carlo approach of Abrams [45]. Instead of starting from an energy function with scaled potentials (as we did in Eq. 2-4), Praprotnik uses force scaling:

$$f_{\alpha\beta} = w(R_\alpha)w(R_\beta)\sum_{i\alpha,j\beta} f^{\mathrm{A}}_{i\alpha j\beta} + [1 - w(R_\alpha)w(R_\beta)]f^{\mathrm{CG}}_{\alpha\beta} \tag{2-17}$$

Here, w are the scaling functions of the interacting beads α and β. This scheme also obeys Newton's third law and is constructed not to cause any flux of particles over the intermediate healing region. Unfortunately, the energy function is not known in this approach. Another difference is the use of fractional degrees of freedom. The temperature is then calculated using the fractional analog of the equipartition theorem

$$< K_\alpha >= \frac{w(R_\alpha)k_B T}{2} \tag{2-18}$$

where $< K_\alpha >$ is the average kinetic energy per fractional degree of freedom [46, 47]. In this approach, all interactions, including the atomistic intra-bead interactions are scaled, and simultaneously the kinetic energy of the atomic degrees of freedom is scaled to zero when a particle switches from atomistic to coarse-grained resolution. Control of the kinetic energy is established through a position dependent dissipative particle dynamics thermostat coupled to each (fractional) degree of freedom.

The other approach worth mentioning is to derive a mixed-resolution Hamiltonian, starting from a linear combination of all possible Lagrangians that can be constructed by treating all particles in the atomistic region plus a subset of those in the healing region at the high-resolution level of theory, and treating all other particles at the low resolution [9]. Also in this case, the internal, high-resolution, degrees of freedom gradually "freeze" when particles leave the atomistic region, through a coordinate dependent kinetic energy. Although this approach is somewhat more involved, requiring the implementation of a special generalized integrator and the definition of a special mixed-resolution potential energy function, one advantage is that it is not limited to pair-potential forcefields but can also be used in combination with many-body forcefields, such as is common in a quantum mechanical treatment [48]. Also, the total momentum and total energy are strictly conserved.

2.3. REVERSE MAPPING THROUGH RIGID BODY ROTATION

From the viewpoint of the reverse mapping problem, the example of liquid methane shown in the previous section represents the simplest possible case. The atomistic methane molecule, CH_4, is symmetrical: it has four hydrogen atoms in a tetrahedral arrangement around a central carbon atom, and hence to a first approximation is well-described by the united atom sphere which is used to represent it in the coarse grained region. When such a united atom sphere enters the healing region, one may

conceivably insert the missing hydrogen atoms in a randomly oriented tetrahedron without paying too steep a price in potential energy. The tetrahedron will need to rotate to avoid steric clashes with nearby methane molecules, and this motion is the source of the energy flow to the thermostat seen in the previous section. These steric clashes are non-bonded in nature, and non-bonded forces are the softest forces which exist in the forcefield. Soft forces are easily "healed" in the healing region. Once one moves from methane to more complex molecules, the reverse mapping problem can become serious, and one needs a more sophisticated strategy to solve it.

Recall that our solution to the reverse mapping problem is to associate a frozen atomistic fragment with each coarse-grain site, and to rigidly rotate these fragments about their COM in accordance with an energy function designed to maintain a low-energy atomistic-level representation of the system. The atomistic fragments are treated as rigid bodies with no internal degrees of freedom. This means that there are only three degrees of freedom per coarse-grain site to evolve, namely an element of the special orthogonal Lie group SO(3). In the remainder of this section we present several strategies for implementing this SO(3) rotational motion. First we present an energy minimization approach using conjugate gradient optimization. Next, we describe a dynamics approach using the direct analog of the velocity Verlet algorithm for the SO(3) setting. We then discuss the coupling between the SO(3) dynamics and the coarse-grained molecular dynamics, and finish this section with a case study of a polyethylene chain.

2.3.1. Rigid Body Rotational Optimization

Here we present an algorithm that uses SO(3) optimization to align molecular fragments corresponding to coarse-grained sites. The output from this algorithm consists of rigid fragments centered at the coarse-grained sites, rotated to minimize an energy function consisting of both intra- and inter-molecular terms. The approach is based on an algorithm due to Taylor and Kriegman [49] in which a sequence of local parameterizations of the manifold SO(3) is used, rather than relying on a single global parameterization such as the Euler angles. The problems caused by singularities in a global parameterization are thus avoided.

One can object to an energy-minimized structure on the grounds that it is not compatible with the ensembles typically used in molecular dynamics simulations (e.g. NVT or NPT). In this sense, one can ask the question "In what sense is this configuration a representation of the underlying molecular system?" Rather than attempting to provide a mapping algorithm that generates an equilibrated atomistic configuration, in this section we take a more pragmatic approach. The mapping algorithm described here quickly generates a globally stable atomistic configuration that further requires very localized relaxation and equilibration.

The algorithm minimizes a real-valued objective function $E{:}SO(3) \to \mathrm{R}$ defined on the set of rotation matrices

$$R \in SO(3) \equiv \left\{ \in \mathrm{R}^{3\times 3} : R^t R = I,\ \det(R) = 1 \right\}. \tag{2-19}$$

At every point R_0 on the manifold $SO(3)$ we construct a continuous, differentiable mapping between a neighborhood of R_0 on the manifold and an open set in R^3,

$$R(\omega) = R_0 \exp J(\omega),\ \omega \in \mathrm{R}^3,\ |\omega| < \pi \tag{2-20}$$

where the skew symmetric operator $J{:}\mathrm{R}^3 \rightarrow SO(3)$ is defined as

$$J(\omega) = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} . \tag{2-21}$$

$R(\omega)$ can be computed using the Rodrigues formula, although we will not need to do so. The objective (energy) function can be expanded to quadratic order about R_0 as

$$E(R(\omega)) = E(R_0) + g^t\omega + \omega^t H\omega \tag{2-22}$$

where g and H are the gradient and the Hessian of the function, respectively, evaluated at the point $\omega = 0$ which corresponds to the rotation matrix R_0. The conjugate gradient incremental step is

$$\omega_s = -H^{-1}g\,. \tag{2-23}$$

This incremental step determines the new rotation matrix as follows:

$$R = R_0 \exp J(\omega_s)\,. \tag{2-24}$$

The incremental step must lie within the range of the local parameterization, namely $|\omega_s| < \pi$. The updating step can be made computationally efficient by representing the rotations by unit quaternions. The relationship between $SO(3)$ and the group of unit quaternions $Sp(1)$ is

$$q = (\cos\theta, \hat{\omega}\sin\theta),\ \theta = |w|/2\,. \tag{2-25}$$

The incremental step corresponds to the quaternion

$$q_s = \left(\cos\frac{\theta}{2}, \omega\frac{\sin(\theta/2)}{\theta}\right) \text{ where } \theta = |w|\,. \tag{2-26}$$

With the rotation R_0 expressed as the unit quaternion q_0, the product of the two rotations, which gives the new rotation matrix, is given by the quaternion multiplication $q_0 q_s$. It has been shown that this algorithm exhibits quadratic convergence provided that the starting point is sufficiently close to a minimum. According to Eq. (2-22), we are supposed to evaluate the Hessian as well as the gradient to compute the update step. However, by employing the Fletcher-Reeves-Polak-Ribiere version of the conjugate gradient algorithm, only the gradient is needed [50].

To apply this algorithm to molecular systems, two things are needed. Firstly, an objective function must be chosen which imbues the algorithm with chemical meaning. This function should provide a measure of the potential energy of the molecular configuration associated with a given rotation matrix. Secondly, the algorithm must be extended to many coupled SO(3) optimizations so that the molecular system is simultaneously and concertedly optimized over all the coarse-grain centers. This multi-body extension is in fact trivial and does not incur any additional computational cost aside from the necessary linear scaling with the number of centers. The nature of the multi-body aspect of the algorithm will become clear in what follows.

Let us now address the choice of an energy function. Only interactions between atoms belonging to different coarse-grain units need be considered. This is because the intra-unit degrees of freedom are frozen. The SO(3) algorithm is designed to find the optimal rotational orientation of each of these fragments, where the center of mass of each fragment is constrained to lie at the location of the coarse-grained site representing it, and where no internal relaxation of the intra-fragment degrees of freedom is allowed. Ideally, we would like to take the functional form and the parameters of all of the terms contributing to the energy function from an underlying atomistic force field. There is no need to invent new potential energy terms when we have well-parameterized ones at our disposal. The most important contribution to the energy function is a bonded term arising from the "dangling" bonds in the molecular fragments which would connect the fragments to one another in an atomistic representation of the system. This function is harmonic in the interatom distance and is expressed as (see Figure 2-4)

$$E(R_1, R_2) = k/2\,(|r + R_2 v - R_1 u| - d_0)^2 \tag{2-27}$$

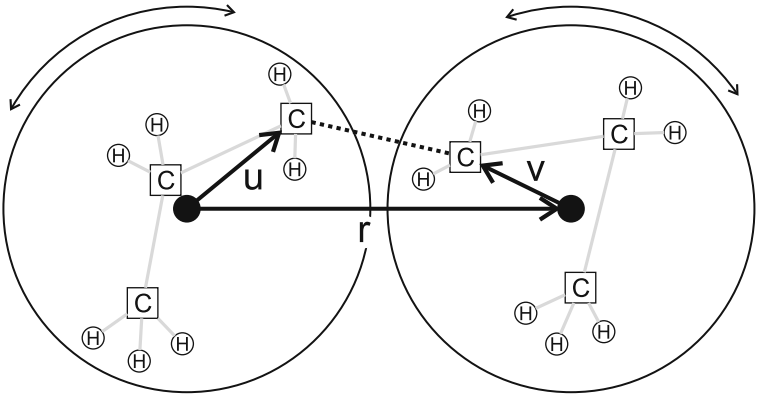


Figure 2-4. Schematic of the SO(3) optimization algorithm for a hexane molecule. Atomistic fragments are rotated about the centers of mass of the coarse-grain units representing them to align the unconnected atomistic bond between fragments (shown in *dashed line*)

where R_1 and R_2 are the rotation matrices corresponding to coarse-grain units 1 and 2, located at positions COM1 and COM2, respectively. The vector from COM1 to COM2 is denoted r. u represents the vector from COM1 to the atom in coarse-grain unit 1 involved in the bond. v represents the vector from COM2 to the atom in coarse-grain unit 2 at the other end of the bond. The bond has an equilibrium distance of d_0 and a force constant of k. In order to perform optimization, the gradient must be evaluated. There are six gradient terms associated with this energy function, three for rotation matrix R_1 and three for R_2. For the ω_x component of R_1, the gradient is

$$\frac{\partial}{\partial \omega_{x1}} E = -k \frac{(|r + R_2 v - R_1 u| - d_0)}{|r + R_2 v - R_1 u|} (r + R_2 v - R_1 u) \cdot \frac{\partial}{\partial \omega_{x1}} R_1 u \tag{2-28}$$

with

$$\frac{\partial}{\partial \omega_{x1}} R_1 u = R_0^1 J(\hat{x}) u \tag{2-29}$$

where R_0^1 denotes the initial rotation matrix for coarse-grained center 1. This last result is computationally important because it means that the J operator only ever acts on three elements, $\hat{x}$, $\hat{y}$, and $\hat{z}$, where,

$$J(\hat{x}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, J(\hat{y}) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, J(\hat{z}) = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2-30}$$

The multi-body nature of the minimization procedure proceeds as follows: the potential energy function is written as a sum over all the coarse-grain sites, with separate terms arising from bonds, bends, torsions, one-fours, and Lennard-Jones interactions. The gradient of this function is evaluated with respect to each degree of freedom, namely the three numbers ω_x, ω_y, and ω_z for each coarse-grain site. This gradient is used to decide upon a global incremental update step in which all the rotation matrices are simultaneously changed. For further details on how to implement bend, torsional, and non-bonded energy terms, we refer the interested reader to our previous work [51].

2.3.2. Rigid Body Rotational Dynamics

In the interest of energy conservation, we now turn to rigid body rotational dynamics algorithms, namely how to adapt the velocity Verlet algorithm for the setting of the SO(3) Lie group. This is a challenging problem because the dynamics occurs on a curved manifold, not a vector space. There are no known algorithms which possess all the desired properties expected of such an integrator, namely that it be symplectic, time-reversible, and energy and momentum conserving. We have chosen to implement the best-performing currently known explicit algorithm for rigid body dynamics, which is due to Krysl and Endres [52]. This algorithm is the direct analog

of the velocity Verlet algorithm for the rotation dynamics setting. The fundamental law of motion for rotational dynamics is $t = \dot{L}$ where t is the torque, $L = I\omega$ is the angular momentum, I is the moment of inertia tensor, and ω is the angular velocity. The initialization and subsequent dynamics components of the algorithm are as follows.

We initialize the angular velocity from the coordinates and (linear) velocities of the atoms via

$$\omega_0 = I_0^{-1} L_0 = I_0^{-1} \left(\sum_i r_i \times m_i v_i \right), \tag{2-31}$$

where I is computed relative to the COM, r_i is the vector from the COM to atom i; m_i is the mass of atom i, v_i is the (linear) velocity of particle i, and $\times$ is the cross product. The angular acceleration is initialized from

$$\alpha_0 = I_0^{-1} \left(t_0 - \omega_0 \times I_0 \omega_0 \right) , \tag{2-32}$$

where t is the torque.

Following initialization, the dynamics occurs in a 4-step algorithm as follows, where the time step is τ.

Step 1: $\omega'_n = \omega_{n-1} + \frac{\tau}{2}\alpha_{n-1}$
Step 2: $R_n = \exp\left[\tau J(\omega'_n)\right]$
Step 3: Solve $I_n\alpha_n - t_n + \left(\omega'_n + \frac{\tau}{2}\alpha_n\right) \times I_n(\omega'_n + \frac{\tau}{2}\alpha_n) = 0$ for α_n
Step 4: $\omega_n = \omega'_n + \frac{\tau}{2}\alpha_n$

Steps 1 and 4 are the velocity half-steps. In Step 2 the rotation matrix R_n is used to update the atomic positions relative to their COM. Hence, unlike the original algorithm, we do not apply incremental updates to the rotation matrix, but instead directly rotate the atoms at each time step. We remind the reader that $J(\omega)$ is defined by Eq. (2-21). In between Steps 2 and 3, the torque and the moment of inertia tensor are updated. In Step 3 we need to solve a non-linear vector equation for the angular acceleration α_n using Newton′s method since it is coupled with the angular velocity due to the velocity update. The Jacobian (the matrix of first partial derivatives) required for Newton′s method is straightforward to evaluate [52].

The multi-body nature of the algorithm, as was the case for the optimization algorithm presented above, is trivial: the net torque on each coarse-grained center is computed due to the effect of all other fragments one wishes to consider.

2.3.3. Coupling Between the Rotational Dynamics and Coarse-Grained Molecular Dynamics

To allow for reverse mapping on the fly during a coarse-grained molecular dynamics simulation, we wish to couple the velocity Verlet algorithm introduced above for the rigid body rotational dynamics to the dynamics of the coarse-grained centers of mass. That is, we wish to perform the rotational dynamics of the atoms within

the coarse-grained beads simultaneously with the (Cartesian) molecular dynamics of the beads in a fashion reminiscent of Car-Parrinello MD. In the latter method, the classical dynamics of the nuclei are in principle coupled to a second, artificial, dynamics of electronic wave function degrees of freedom, although in that case an adiabatic separation exists due to the large difference in particle masses, and therefore in temperatures, between the two sub-systems. Here, such a separation does not exist.

The orientation of the rigid body and its rotational motion are governed through atomistic interactions that span different beads as illustrated by the dashed C–C bond in Figure 2-4. These atomistic interactions are affected if we allow the beads to move with respect to each other. Imagine for example that the two beads in Figure 2-4 are moving away from each other, than clearly also the (dashed) C–C bond elongates, leading to an increased torque on the atomistic body and thus a speedup of the rotational motion. That means in practise that the rotational dynamics continuously heats up if we allow the beads to move, unless a corresponding back-coupling of the rigid body rotation is added to the dynamics of the beads or, alternatively, a friction is added to the rotational dynamics to avoid heating up. The latter damped dynamics is most easily implemented and results in an alternative on-the-fly rotational optimization scheme (see also Section 2.3.1); this solution is an example of a general technique known as simulated annealing. Here instead, we will discuss the (back-) coupling between the (undamped) dynamics subsystems, as it will play a role in the hybrid multiscale method (see Section 2.2).

Adding the correct back-coupling of the rotational dynamics to the dynamics of the beads entails adding the atomistic interactions that govern the rotational motions to the coarse-grained dynamics. However, we do not wish to disrupt the coarse-grained molecular dynamics more than necessary, not in the least because adding stiff atomistic interactions to the coarse-grained dynamics would require a smaller time step for its evolution. We therefore distinguish between translational and rotational motion of the beads with respect to each other, where we define translation as the motion that alters the distance between two beads and rotation as the motion in which the bead-bead distance remains constant.

For the translational motion of beads (imagine the coarse-grained stretch vibration between the two beads in Figure 2-4), instead of adding the back-coupling of the atomistic interaction to the coarse-grained dynamics, we remove the coupling of the coarse-grained translational motion to the rotational dynamics. This is done by re-normalizing the atomistic configurations to the equilibrium bead-bead distance, instead of using the actual bead-bead distance. That is, the instantaneous bead-bead distance r in Eq. (2-27) for the rotational potential energy is replaced by the equilibrium bond distance r_0, as follows

$$E(R_1, R_2) = k/2\,(|r_0 + R_2 v - R_1 u| - d_0)^2 \tag{2-33}$$

and likewise in the force expression of Eq. (2-28). This way, the rotation dynamics is computed as if the beads are always placed at their equilibrium distance, thus

removing any effect of translational motion on the rotational dynamics and the need to add back-coupling.

For the rotational motion of the beads, we cannot remove the coupling because it would defeat the purpose of keeping the atomistic configuration while the (coarse-grained) molecule rotates. That is, when the coarse-grained molecule rotates, we want the atomistic rigid bodies to adapt their orientation simultaneously. The fluctuations on the interactions due to the rotation of the molecule are expected to be much smaller than that due to the coarse-grained stretching modes. The proper back-coupling is added to the coarse-grained dynamics using

$$f_1 \times r = T_1 = f_i \times u \tag{2-34}$$

That is, a back-coupling force f_1 on bead 1 is computed from the torque T_1 on the atomistic rigid body of this bead that is due to the atomistic interaction between one of its atoms i and another atom j belonging to neighboring bead 2 (see Figure 2-4) . In other words, the torque from the force f_i on atom i that drives the rotation results in addition to a force on the bead that is normal to the plane of the torque and the bond, r, between the beads.

Having added the proper coupling between the rotational rigid atom dynamics and the rotational motions of the coarse-grained molecules (and removed the coupling with the translational bead motions) we obtain a stable energy conserving dynamics scheme which tends to thermal equilibrium between the two subsystems. In the following section, we will compare the two methods of on-the-fly reverse mapping through rigid rotation, and show that both the optimization scheme (or simulated annealing scheme) as well as the coupled dynamics scheme succeed in recovering good approximations of the atomistic structure in a coarse-grained simulation.

2.3.4. Case Study 2: Polyethylene Chain

In this case study, we compare the reverse mapping schemes, introduced in the previous section, for coarse-grained simulations of a ($C_{75}H_{152}$) polyethylene chain in vacuum. The chain is represented by 25 coarse-grained beads, interacting through harmonic bond and bending potentials and Lennard–Jones type van der Waals interactions [53]. Each bead is mapped on 3 carbon atoms and its associated hydrogens as shown in Figure 2-5. We performed four NVT coarse-grained dynamics simulations of 500 ps in length at $T = 303$ K, each with a different variant of the reverse mapping scheme to recover atomistic configurations on the fly. The first 100 ps are regarded as equilibration; the remaining 400 ps of the trajectory was analyzed and compared with that of a fully atomistic simulation of the system. The four reverse mapping schemes are:

1. At each timestep the relaxed atomistic C_3H_6 structure is inserted into each bead with a random orientation (i.e. random mapping).

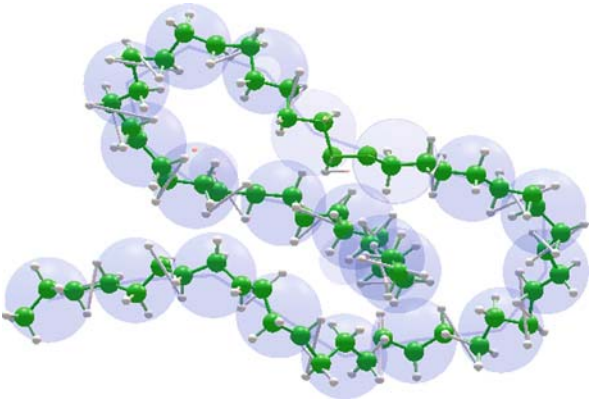


Figure 2-5. Illustrative snapshot from a coarse-grained simulation of a polyethylene chain ($C_{75}H_{152}$), using the reverse mapping through rigid body rotation to recover atomistic configurations. Rotational dynamics of the atomistic C_3H_6 functions is governed by a simplified energy function containing only a harmonic C–C bond plus two repulsive H–H bonds (shown in *white*) between neighboring beads

2. The C_3H_6 structures are inserted at the beginning after which their orientations are optimized during the coarse-grained simulation through the damped SO(3) rotational dynamics (i.e. damped rigid rotation).
3. Same as 2. but instead of annealing, the rotational dynamics is used, coupled to the coarse-grained dynamics (i.e. coupled rigid rotation).
4. atomistic flexible rotation

Here, the first scheme of random orientations is used for reference, the second and third schemes are the two rigid body rotation schemes introduced in the previous section, and the fourth scheme is another rotational scheme that acts on flexible atomistic structures, rather than rigid bodies.

This last flexible scheme is added here because we will use it later in the hybrid multiscale method (see Section 2.2). That is, until now we have talked about reverse mapping to recover atomistic configurations from a coarse-grained trajectory, but hereafter, these reverse mapping schemes will be combined with the hybrid MD algorithm to pre-condition coarse-grained molecules before they enter the healing region. In order to gradually switch off this rotational dynamics while the atomistic interactions are being switched on, we require a variant of the rotational dynamics that will work on flexible, atomistically propagated, bodies as is the case in the healing region. This is the fourth scheme of which the details follow in Section 2.4.

For the purpose of illustration, we used a very simple energy function in all of the three rotational dynamics schemes. That is, in principle we can use the full set of atomistic interactions of bonds, bends, torsions and even non-bonded potentials spanning neighboring beads, to govern the rotational dynamics, but it makes sense to choose an economic reduced subset instead. For our polyethylene chain, an intuitive good choice would include (at least) the carbon-carbon bond interaction spanning each pair of neighboring beads supplemented with the dihedral angles centered at

this C–C bond. Here, in fact, we replace the dihedral angles with two repulsive harmonic interactions between two hydrogens per pair of neighboring beads. These repulsive "bonds" between hydrogens are shown in Figure 2-5 as white sticks.

How do the reverse mapping schemes based on rotational dynamics of rigid or flexible atomistic bodies perform with respect to the randomly oriented bodies and with respect to a fully atomistic molecular dynamics simulation? To answer this question, we compared atomistic distribution functions, two of which are shown in Figure 2-6. On the left-hand side are the histograms of the pair-correlation of each carbon in the chain with the carbon three positions away. This 1–4 correlation is the shortest correlation that always spans two different beads. In the atomistic simulation, the 1–4 correlation shows two peaks reflecting the staggered trans configuration (larger peak) and the two, less favorable, staggered cis configurations (smaller peak) of each quartet of carbons. In the coarse-grained simulations, the repulsive hydrogen interactions (namely the white sticks in Figure 2-5) make the cis configuration very improbable, so that the 1–4 correlations all show a single peak positioned at the larger peak from the atomistic simulation. The broadening of the peak from the reverse mapping schemes is due to the stretch vibration between the beads, which is much softer than that between two carbons. The right-hand panel in Figure 2-6 shows the histograms of dihedral angles between each sequence of 4 sequential carbon atoms centered at the bond between two beads (i.e. the first two carbons belong to one bead and the second two carbons belong to the next bead). Again, the repulsive terms in our simplified energy function used for the rotational dynamics represses completely the secondary peaks seen in the histograms from the atomistic simulation; however all three reverse mapping schemes recover a good approximation to the average structure, which is the trans configuration. As a side remark, we note that the coupled dynamics scheme performs slightly better, showing narrower distributions, than the damped dynamics scheme, which is somewhat surprising as the damped dynamics would be expected to

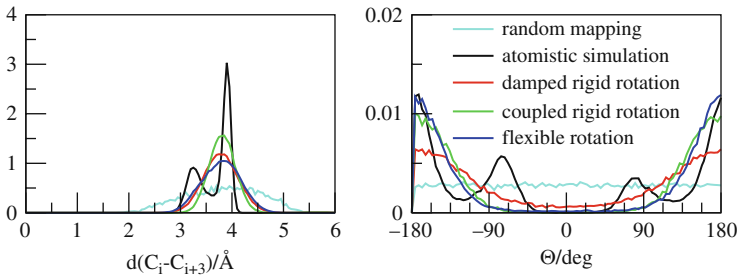


Figure 2-6. Comparison of the reverse mapping schemes to recover the atomistic structure in coarse-grained simulation of a polyethylene chain. *Left panel*: histograms of the distance between each carbon at position i with that at position $i+3$ in the chain. *Right panel*: histograms of dihedral angles of sequential carbon atoms of which the first two carbons and the second two carbons belong to different beads, respectively

tend closer to the lowest energy conformation. Further study to optimize the damping factor might improve this.

All three reverse mapping schemes, based on rotational dynamics of rigid and flexible atomistic bodies, manage very well in maintaining good atomistic configurations on the fly during a coarse-grained simulation. The overhead of the rigid body schemes is minimal due to the local nature of the energy function, relative to the demanding evaluation of the non-bonded interactions in the coarse-grained dynamics. The reverse mapping scheme was initially designed to allow for easy back-and-forth switching between atomistic and coarse-grained simulations of a system, reducing the requirement of re-equilibration in between. The reverse mapping is therefore also expected to be helpful in replica exchange simulations, in which simultaneously several replicas of the system at different resolution are simulated that can exchange based on a Metropolis criterion [18–21]. In the context of this chapter however, we will now return to the hybrid multiscale method and show how the rotational reverse mapping algorithm, applied to the coarse-grained region, is essential to simulate more complex structures.

2.4. COMBINING ROTATIONAL REVERSE MAPPING WITH HYBRID MD

In this last methodological section, we merge the reverse mapping algorithm with the adaptive multiscale method. The recent literature until now only shows applications of adaptive hybrid atomistic/coarse-grained methods that consist of simple spherical beads in the low-resolution representation, for example, methane [7], water [54, 55], and even simpler toy models [8, 56]. The reason for this is that these spherical, or united, atoms require minor re-orientation when switching to the atomistic resolution compared to larger molecules spanning more than one coarse grained bead. In other words, all previous studies have explicitly avoided the reverse mapping problem. Indeed, it is clear from figure 4 that the dashed C–C bond imposes severe constraints on the relative orientation of the two beads representing a hexane molecule. If the atomic fragments in these two beads are randomly oriented with respect to one another when the hexane molecule enters the healing region, a very large healing region will be needed to bring the atomic representation down from an almost infinitely high potential energy value.

Applying the reverse mapping algorithm in the coarse-grained region of the multiscale setup allows, for the first time, hybrid MD simulations of realistic systems with multi-bead molecules in the low-resolution representation. The rigid body rotational dynamics (introduced in Section 2.3.2) maintains rigid atomistic structures superimposed on the coarse-grained molecular dynamics. These rigid structures are rotated about their centers of mass in accordance with an energy function which includes a subset of the local atomistic interactions, such as bond and bending potentials between atoms belong to sequential beads. This pre-conditioning of atomistic configurations in the coarse-grained region therefore requires a much smaller healing region to equilibrate coarse-grained molecules that switch to the atomistic resolution.

Note, however, that applying this reverse mapping scheme only to the coarse-grained region is not enough. Also at the coarse-grained end of the healing region, the orientation of atomistic structures within the beads becomes randomized when the molecules spend sufficient time at healing region positions of, say, around ninety percent coarse-grained character. Only a very wide healing region would then be able to restore the atomistic structure in a smooth manner, when such a molecule is pushed toward the atomistic region and the highly unfavorable atomistic interactions are switched on. This is the reason that a scheme in which atomistic structures are locally equilibrated (constrained to keep the centers of mass at the coarse-grained bead positions), for molecules that leave the coarse-grained region, does not suffice. Instead, we will also apply the rotational reverse mapping scheme in the healing region, which is then gradually switched off, together with the coarse-grained interactions, as the atomistic forcefield takes over.

In the healing region, in contrast to the coarse grained region, the atomistic bodies are not rigid. Flexible bodies break the SO(3) rotational dynamics integrator because the moment of inertia tensor changes due to non-rotational forces. To nevertheless allow for a rotational dynamics in conjunction with the atomistic dynamics, we employ a modified approach that adds the same reduced energy function that governs the rigid body rotational dynamics in the coarse-grained region. However, instead of computing a torque on the body from the atomistic interactions, we let the interaction act directly on the atoms with the additional constraint that the fragment should not feel an effective force. This constraint is satisfied if we apply a counter interaction on the fragment that cancels any translational acceleration and only keeps the rotational acceleration. In the example of Figure 2-4, this would mean that the atomistic bond interaction (dashed line) causes a force on the interacting carbon atoms and simultaneously a constraint force of the same amplitude but with opposite sign on the centers of mass. In the healing region, the coarse-grained forces acting on the centers of mass are distributed, mass-weighted, over the atoms. The coupling of this rotational dynamics with the normal atomistic dynamics in the healing region is implemented on the same footing as the coupling between those terms in the coarse-grained region (see Section 2.3.4). In Section , this atomistic flexible rotation variant of the reverse mapping technique was shown to behave well, compared to the rigid body rotational algorithms and to a fully atomistic molecular dynamics simulation.

2.4.1. Case Study 3: Hybrid Simulation of a Polyethylene Chain

In the third case study, we examine the application of the final combined adaptive multiscale molecular dynamics, i.e. including the reverse mapping algorithm, to the folding of a polyethylene chain in vacuum. The $C_{150}H_{302}$ chain is twice as long as in the previous case study and is represented by 50 beads at the coarse-grained level [53]. The initially stretched configuration is taken from a equilibration run, subject to an end-to-end distance constraint. As an illustration of the method, we show 2 short, 150 ps, hybrid MD simulations of the folding process, once using the reverse mapping algorithm and once without. Two atomistic regions are defined

with a radius of $R_A = 6$ Å centered on beads 15 and 30 in the coarse-grained representation of the chains. The surrounding healing region skin is $R_{HR} = 5$ Å. Two representative snapshots of the hybrid MD simulation (using the reverse mapping algorithm) are shown in Figure 2-7.

Comparing the total energies of the two simulations, with and without reverse mapping, displays a dramatic difference, as shown in Figure 2-8, top panel. Using the reverse mapping algorithm, which maintains a good pre-conditioned atomistic structure in the coarse-grained region, the simulation shows very good energy conservation. Without the reverse mapping algorithm the total energy shows erratic behavior (note the difference in scales in these plots), indicating problems and poor accuracy in the simulation. Visual inspection of the trajectory shows that groups of atoms move suspiciously fast in the healing region and eventually, after about 95 ps of simulation, the system explodes. Because of the relatively large number of atoms in the healing region, the increasing temperature (despite the thermostat) is an indication of problems (data not shown). Note that for example in a simulation of the chain in a solvent, such local temperature changes may not be apparent in the total system temperature.

The bottom panel in Figure 2-8 shows the decreasing radius of gyration and the end-to-end distance of the polymer chain as it collapses from the initial stretched configuration to its random coil state. Note that, in this case, the unstable hybrid simulation displays reasonable behavior for these observables, not indicating any obvious problems, at least until the system explodes after 95 ps.

In conclusion, we have seen that incorporating the reverse mapping algorithm into our hybrid MD method to pre-condition the atomistic structure results in a robust adaptive multiscale molecular dynamics method. Close observation of the

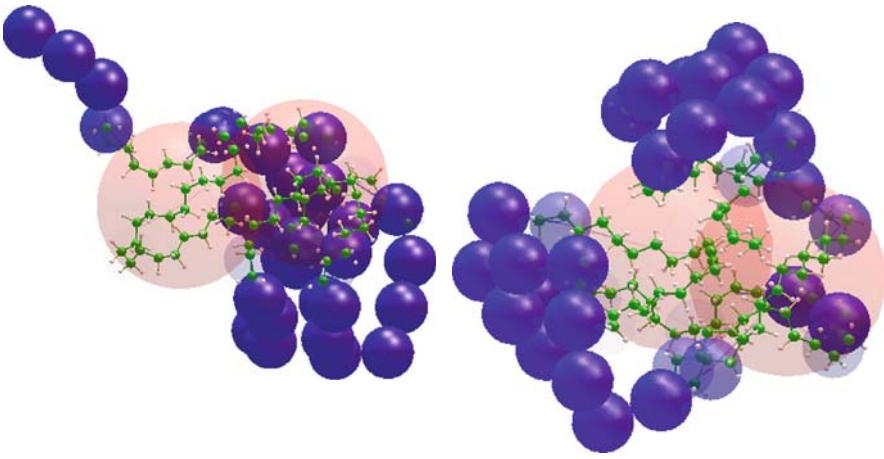


Figure 2-7. Two snapshots from the adaptive multiscale simulation of a polyethylene chain ($C_{150}H_{302}$) in vacuum, using two spherical atomistic regions (*red spheres*). The atomistic regions have a diameter of 6 Å and are centered on beads 15 and 30 of the 50 bead coarse-grained representation (*blue spheres*)

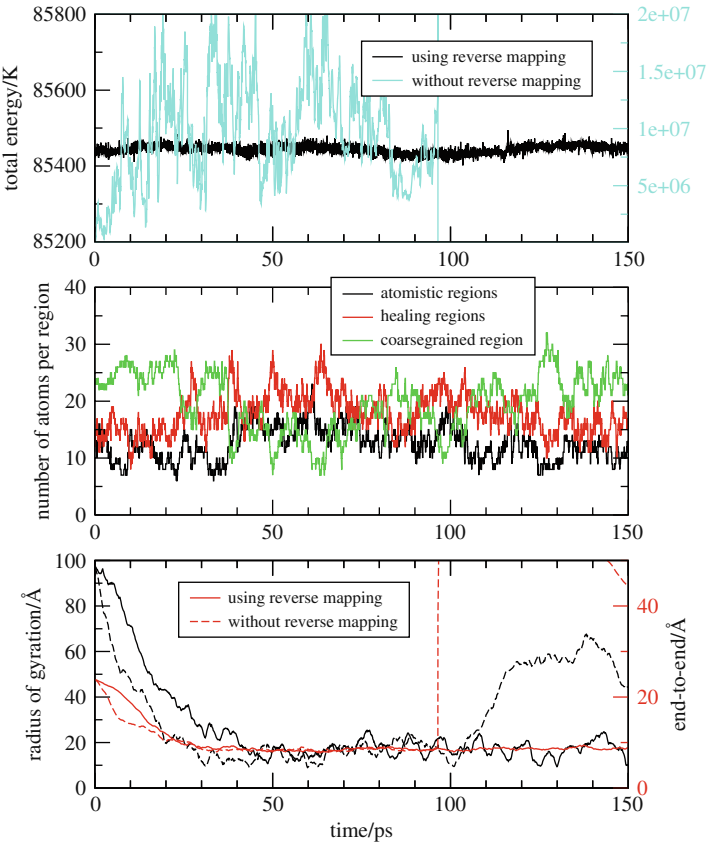


Figure 2-8. Top panel: Using the reverse mapping algorithm within the hybrid MD simulation of a polyethylene chain leads to good conservation of the total energy; instead without reverse mapping the hybrid method is unstable. *Middle panel*: the continuously fluctuating numbers of beads in the low, high, and hybrid resolution regions. *Bottom panel*: the folding of the chain shown by the radius of gyration (*black lines; left axis*) and the end-to-end distance (*red lines; right axis*). Despite the spurious dynamics of the simulation without reverse mapping, these observables show fortuitous similar physical behavior, that is, until the system explodes after about 95 ps

conservation of the total energy is required to assess the accuracy and physical relevance of the hybrid simulation. Without the reverse mapping algorithm, the hybrid method only works for small structureless molecules that can be represented by a single spherical bead in the coarse-grained region.

This case study of the collapse of a polymer chain illustrates the promising potential of the adaptive atomistic/coarse-grained method for the study of much more complicated and demanding phenomena of self-assembly, for example protein-protein interaction, multimeric protein assembly, and protein-DNA interaction. We foresee that such investigations, which now require either enormous computer

resources or rather simplified models, will take full advantage of the speedup from this multiscale method. Compared to a million atom sized MD simulation, such a speedup could easily reach two or three orders of magnitude when the demanding atomistic description can be limited to interacting regions of several hundreds to thousands of atoms.

2.5. SUMMARY

Adaptive multiscale molecular dynamics is a promising new simulation technique aimed at bridging the gap between the large spatial and temporal scales exhibited by the phenomena that we wish to predict and the fast and short-ranged molecular fluctuations that limit current high-resolution methods. This technique allows one to focus the available computational resources on those special regions of the system where the key events are occurring by modeling the system in these regions at a higher, more accurate, resolution than the rest of the system. Here, we used an atomistic description in the high-resolution regions, and coarse-grained models, in which atoms are grouped into single interaction sites, to describe the other, low-resolution, regions.

The level of description of molecules that move between the high and low resolution regions adapts on the fly. This transformation from a coarse-grained description to an atomistic one, or vice versa, proceeds in a smooth manner through a thin intermediate *healing region* that bridges between the high and low resolution regions. In particular, the transformation from coarse-grained to atomistic, the so-called *reverse mapping*, is otherwise cumbersome as it requires the introduction of internal degrees of freedom that should be equilibrated together with their surrounding molecules. The continuous introduction (and removal) of degrees of freedom in the healing region is therefore a non-equilibrium process that produces heat, which can be removed with a thermostat. An important feature of the current multiscale algorithm is that it recovers, nevertheless, the total energy as a conserved quantity. *Energy conservation* provides a crucial handle to assess the accuracy of the integration of the equations of motion (i.e. the quality of the simulation) and is for example necessary to be able to choose an appropriate size for the healing region as well as for the time step.

We also discussed a special reverse mapping technique that allows one to obtain the atomistic conformations from a coarse-grained molecular dynamics simulation at low computational overhead. The algorithm consists of a dynamics on the Lie group SO(3) of rotations for every coarse-grained site. Combining this technique with the adaptive multiscale approach to pre-condition molecules in the low-resolution region, before they enter the healing region, helps to limit the size of the healing region.

ACKNOWLEDGMENTS

We are very grateful to Preston B. Moore, Peter G. Bolhuis, Michael L. Klein and Michele Parrinello for helpful discussions.

REFERENCES

1. Wesolowski, T. A., Warshel, A., *J. Phys. Chem.* **1993**, *97*, 8050.
2. Jacob, C., Neugebauer, J., Visscher, L., *J. Comput. Chem.* **2008**, *29*, 1011.
3. Cramer, C. J., Truhlar, D. G., *Chem. Rev.* **1999**, *99*, 2161.
4. Tomasi, J., Mennucci, B., Cammi, R., *Chem. Rev.* **2005**, *105*, 2999.
5. Tomasi, J., Persico, M., *Chem. Rev.* **1994**, *94*, 2027.
6. Gao, J., in *Reviews in Computational Chemistry*, edited by Lipkowitz, K., Boyd, D., Volume 7, pp. 119–185, 119, VCH, New York, **1992**.
7. Ensing, B., Nielsen, S. O., Moore, P. B., Klein, M. L., Parrinello, M., *J. Chem. Theory Comput.* **2007**, *3*, 1100.
8. Praprotnik, M., Site, L. D., Kremer, K., *J. Chem. Phys.* **2005**, *123*, 224106.
9. Heyden, A., Truhlar, D. G., *J. Chem. Theory Comput.* **2008**, *4*, 217.
10. Shi, Q., Izvekov, S., Voth, G. A., *J. Phys. Chem. B* **2006**, *110*, 15045.
11. Neri, M., Anselmi, C., Cascella, M., Maritan, A., Carloni, P., *Phys. Rev. Lett.* **2005**, *95*, 218102.
12. Villa, E., Balaeff, A., Mahadevan, L., Schulten, K., *Multiscale Model. Simul.* **2004**, *2*, 527.
13. Dupuy, L. M., Tadmor, E. B., Miller, R., Phillips, R., *Phys. Rev. Lett.* **2005**, *95*, 060202.
14. Diestler, D. J., Zhou, H., Feng, R., Zeng, X. C., *J. Chem. Phys.* **2006**, *125*, 064705.
15. Rudd, R. E., Broughton, J. Q., *Phys. Stat. Sol. B* **2000**, *217*, 251.
16. Izvekov, S., Voth, G. A., *J. Phys. Chem. B* **2005**, *109*, 2469.
17. Liu, P., Izvekov, S., Voth, G. A., *J. Phys. Chem. B* **2007**, *111*, 11566.
18. Lyman, E., Ytreberg, F., Zuckerman, D., *Phys. Rev. Lett.* **2006**, *96*, 028105.
19. Lyman, E., Zuckerman, D. M., *J. Chem. Theory Comput.* **2006**, *2*, 656.
20. Christen, M., van Gunsteren, W., *J. Chem. Phys.* **2006**, *124*, 154106.
21. Liu, P., Voth, G. A., *J. Chem. Phys.* **2007**, *126*, 045106.
22. Csányi, G., Albaret, T., Payne, M. C., DeVita, A., *Phys. Rev. Lett.* **2004**, *93*, 175503.
23. Hoogerbrugge, P. J., Koelman, J. M. V. A., *Europhys. Lett.* **1992**, *19*, 155.
24. Smith, D. E., Harris, C. B., *J. Chem. Phys.* **1990**, *92*, 1304.
25. Turq, P., Lantelme, F., Friedman, H. L., *J. Chem. Phys.* **1977**, *66*, 3039.
26. Laio, A., Parrinello, M., *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562.
27. Ensing, B., Vivo, M. D., Liu, Z. W., Moore, P. B., Klein, M. L., *Acc. Chem. Res.* **2006**, *39*, 73.
28. Nielsen, S. O., Lopez, C. F., Srinivas, G., Klein, M. L., *J. Phys. Condens. Mater.* **2004**, *16*, R481.
29. Henderson, R. L., *Phys. Lett. A* **1974**, *49A*, 197.
30. Lyubartsev, A. P., Laaksonen, A., *Phys. Rev. E* **1995**, *52*, 3730.
31. Soper, A. K., *Chem. Phys.* **1996**, *202*, 295.
32. Akkermans, R., Briels, W., *J. Chem . Phys.* **2001**, *114*, 1020.
33. Jain, S., Garde, S., Kumar, S. K., *Ind. Eng. Chem. Res.* **2006**, *45*, 5614.
34. Njo, S. L., van Gunsteren, W. F., Mueller-Plathe, F., *J. Chem. Phys.* **1995**, *102*, 6199.
35. Ercolessi, F., Adams, J. B., *Europhys. Lett.* **1994**, *26*, 583.
36. Izvekov, S., Parrinello, M., Burnham, C. J., Voth, G. A., *J. Chem. Phys.* **2004**, *120*, 10896.
37. Maurer, P., Laio, A., Hugosson, H. W., Colombo, M. C., Rothlisberger, U., *J. Chem. Theory Comput.* **2007**, *3*, 628.
38. de Pablo, J. J., Curtin, W. A., *MRS Bull.* **2007**, *32*, 905.
39. Nosé, S. J., *J. Chem. Phys.* **1984**, *81*, 511.
40. Hoover, W. G., *Phys. Rev. A* **1985**, *31*, 1695.
41. Parrinello, M., Rahman, A., *Phys. Rev. Lett.* **1980**, *45*, 1196.
42. Bussi, G., Donadio, D., Parrinello, M., *J. Chem. Phys.* **2007**, *126*, 014101.

43. MacKerel Jr., A., Brooks III, C., Nilsson, L., Roux, B., Won, Y., Karplus, M., CHARMM: The energy function and its parameterization with an overview of the program, in *The Encyclopedia of Computational Chemistry*, edited by v. R. Schleyer et al., P., Volume 1, pp. 271–277, John Wiley & Sons: Chichester, 1998.
44. Jorgensen, W., Madura, J., Swenson, C., *J. Am. Chem. Soc.* **1984**, *106*, 6638.
45. Abrams, C. F., *J. Chem. Phys.* **2005**, *123*, 234101.
46. Praprotnik, M., Kremer, K., Site, L. D., *J. Phys. A Math. Theor.* **2007**, *40*, F281.
47. Praprotnik, M., Kremer, K., Site, L. D., *Phys. Rev. E* **2007**, *75*, 017701.
48. Heyden, A., Lin, H., Truhlar, D. G., *J. Phys. Chem. B* **2007**, *111*, 2231.
49. Taylor, C. J., Kriegman, D. J., *IEEE Trans. Rob. Autom.* **1998**, *14*, 417.
50. Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., *Numerical Recipes*, Cambridge University Press: New York, **1992**.
51. Nielsen, S. O., Ensing, B., Moore, P. B., Klein, M. L., Coarse grained to atomistic mapping algorithm: a tool for multiscale simulations, in *Multiscale simulation methods for nanomaterials*, edited by Ross, R., Mohanty, S., pp. 73–88, John Wiley and Sons, Inc.; Hoboken, NJ, **2008**.
52. Krysl, P., Endres, L., *Int. J. Numer. Methods Eng.* **2005**, *62*, 2154.
53. Nielsen, S. O., Lopez, C. F., Srinivas, G., Klein, M. L., *J. Chem. Phys.* **2003**, *119*, 7043.
54. Praprotnik, M., Matysiak, S., Site, L. D., Kremer, K., Clementi, C., *J. Phys. Condens. Mater.* **2007**, *19*, 292201.
55. Matysiak, S., Clementi, C., Praprotnik, M., Kremer, K., Site, L. D., *J. Chem. Phys.* **2008**, *128*, 024503.
56. Praprotnik, M., Site, L. D., Kremer, K., *Phys. Rev. E* **2006**, *73*, 066701.

CHAPTER 3

TRANSITION PATH SAMPLING STUDIES OF SOLID-SOLID TRANSFORMATIONS IN NANOCRYSTALS UNDER PRESSURE

MICHAEL GRÜNWALD AND CHRISTOPH DELLAGO
University of Vienna, Faculty of Physics, Boltzmanngasse 5, 1090, Vienna, Austria
e-mail: michael.gruenwald@univie.ac.at; christoph.dellago@univie.ac.at

Abstract: Many interesting phenomena in nature, as diverse as first order phase transformations, biomolecular isomerizations, or transport processes in solids, are characterized by widely disparate timescales. While the waiting time for a spontaneous incidence of such a process can exceed seconds or even hours, the underlying relevant molecular motions occur on the femtosecond timescale. This fact poses a serious problem to molecular dynamics computer simulations aimed at revealing the atomistic mechanisms of such phenomena. Here, we give a review of transition path sampling, a set of computational methods designed to overcome this timescale problem. As an application, we show how transition path sampling can be used to identify the atomistic mechanisms of structural transformations in nanocrystals under pressure

Keywords: Rare events, Transition path sampling, Nanoparticles

3.1. RARE EVENTS IN COMPUTER SIMULATIONS

In the past few decades, molecular dynamics simulation has grown into a very powerful tool that today is used routinely to study the dynamics of condensed matter systems consisting of up to a few million particles with atomistic resolution. Many processes occurring in nature and technology such as the folding of a protein or the transport of a dopant through a semiconductor, however, are still beyond the reach of this methodology due to widely disparate time scales that are present in the problem. Consider, for instance, the nucleation of a crystal from the undercooled liquid. For moderate undercooling, this process typically proceeds through the formation of a critical nucleus that then grows, eventually transforming the whole sample into the crystalline state. Since this process involves the creation of an interface between the crystallite and the metastable liquid, which is associated with a free energetic barrier, the formation of the critical nucleus is rare on the time scale of basic

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 61–84.
DOI 10.1007/978-1-4020-9785-0_3,

molecular motions. Indeed, it has been known for a long time that water, carefully cooled below the freezing point, can remain in this supercooled state for hours or even days. Thus, the time scale for nucleation exceeds the picosecond time scale for the formation and cleavage of hydrogen bonds by many orders of magnitude. Similar rare but important events, related to high energy barriers or entropic bottlenecks in phase space, can also dominate the dynamics of folding proteins, chemical reactions or transport on surfaces.

Naturally, such a wide separation of time scales is a problem for molecular dynamics simulation. In this method, the equations of motion of the system are solved numerically in small time steps. The size of the time step must be selected such that even the fastest motions in the system are reproduced faithfully. In a molecular system, fast bond and angular vibrations require a time step of about 1 femtosecond. With such a time step, current computer technology permits to follow the time evolution of the system for $10^6 - 10^9$ time steps, corresponding to total simulation times from nanoseconds to microseconds. Of course, the accessible simulation times depend on the size of the system and on the particular way of calculating the forces acting on the individual atoms. If forces are determined ab initio by solving of the electronic structure problem, typical simulation times do not exceed tens of picoseconds even for moderate system sizes of 100–200 atoms. For our example, the formation of a crystal from the supercooled liquid, this limitation in the accessible time scales means that in a molecular dynamics simulation the crystallization event simply cannot be observed.

For the computer simulator this situation is frustrating, particularly because typically rare events are not *slow*. Rather, if they occur, they occur rapidly. For instance, the formation of a critical crystalline nucleus proceeds quickly, while the time spent waiting for this event may be very long. (In fact, microscopic time reversibility requires that the formation of a critical nucleus happens as quickly as its decay.) Similarly, an activated chemical reaction can proceed quickly once it is initiated, but the waiting time between subsequent reactions may be very long. To circumvent this problem caused by widely disparate time scales, several computer simulation algorithms have been devised in recent years. If the reaction mechanism is known in terms of a reaction coordinate that quantifies the reaction progress, for instance the size of the crystalline nucleus forming in the supercooled liquid, the rare event can be studied with umbrella sampling [1] or the blue moon sampling technique [2]. In these methods, an appropriate bias or constraint forces the system to visit the configurations associated with rare barrier crossing events. The detailed mechanism and rate constants of the transition can then be studied using the Bennett-Chandler approach [3, 4], in which dynamical trajectories are initiated from these rare configurations, expanding on the original idea of transition state theory [5, 6].

In complex molecular systems, however, a priori knowledge of the reaction mechanism is often not available and these methods are not directly applicable. In such cases, methods that modify the dynamics such as metadynamics [7], temperature accelerated dynamics [8], or coarse molecular dynamics [9] can be used to explore the possible mechanisms for transitions between stable states. If both the

initial and the final state are known, the transition path sampling (TPS) method, an importance sampling scheme acting in trajectory space, can be used to study the transition. In contrast to other methods, in a transition path sampling simulation truly dynamical trajectories are considered such that both the mechanism as well as the kinetics of the transition can be determined. Alternative methods to study rare transitions between known (meta)stable states are the nudged elastic band (NEB) method [10] and the string method [11, 12]. For a recent review of these approaches and a discussion of their relation to the transition path sampling methodology we refer the reader to [13]. In the present article, we will concentrate on the transition path sampling method and its application.

A typical example to which transition path sampling can be fruitfully applied are phase transformation occurring in semiconductor nanocrystals under pressure. At ambient pressure, cadmium selenide exists in the wurtzite structure, in which every sadmium atom is coordinated by exactly four selenium atoms. Under pressure, this material undergoes a transformation to the denser rocksalt structure, in which every atom is coordinated by six atoms of the other species. If one goes from the bulk system to small crystallites of nanoscopic dimensions, the pressure at which the transition occurs increases considerably [14]. This finite size effect can be understood qualitatively in terms of the higher surface free energy of the rocksalt structure. In small crystals, this surface free energy plays an important role such that, in comparison to the bulk, a larger pressure has to applied to the system to compensate for the extra free energetic cost of the surface. It is conceivable that transitions in the bulk and in the nano-crystal differ not only by the pressure at which they occur, but also by the specific mechanism that transforms one phase into the other. The mechanism preferred in the bulk may be blocked in the nano-crystal, because it leads to morphologies with particularly unfavorable surface free energetics. While experiments have yielded detailed information on the thermodynamics and kinetics of this transition in CdSe nanocrystals, their temporal and spatial resolution is not sufficient to follow the atomic motions during the transition and reveal the mechanism. In principle, molecular dynamics simulations can provide this information. It turns out, however, that in order to observe the transition in a straightforward molecular dynamics simulation within the accessible simulation time, the pressure has to exceed the pressure applied in the experiments by far. At more realistic pressures, the typical waiting times are simply too long for a transformation to be observed in such a simulation. This time scale problem can be solved using transition path sampling [15–17]. In this method, one concentrates on those segments of the time evolution in which the transition event actually happens. An analysis of the trajectories harvested with transition path sampling then yields both the mechanism and the rate of the transition. In the following, we will first outline the main concepts of transition path sampling and then explain how this computational technique can be used to study phase transformations occurring in nanocrystals under pressure.

The remainder of this article is organized as follows. In Section 3.2 we briefly review the main ideas and algorithms of transition path sampling including methods for the analysis of transition pathways as well as for the calculation of reaction rate constants. A specific variant of the transition path sampling algorithm designed

for the simulation of pressure induced structural phase transitions in nanocrystals is then discussed in Section 3.3. The application of this algorithm to the Wurtzite-to-Rocksalt transition of CdSe nanocrystals is the subject of Section 3.4. A few concluding remarks are offered in Section 3.5.

3.2. TRANSITION PATH SAMPLING

Transition path sampling is a computational methodology developed to study rare transitions between long-lived metastable states [18, 19]. These stable states, let us call them A and B, can be different phases of a condensed material in the case of phase transitions or different chemical species in the case of a chemical reaction. Transitions between stable states A and B are rare (otherwise we could study them with standard molecular dynamics simulation) and may involve crossings of possibly rough and unspecified free energy barriers. While transition path sampling does not require an prior knowledge of the transition mechanism, the stable states A and B between which the transition occurs must be known in advance. The central idea of transition path sampling now is to consider only short trajectories, long enough for the barrier crossing event to complete, but much shorter than the typical waiting time before transitions occur. These short trajectories have different probabilities to be observed: trajectories fluctuating in the stable states, for instance, are more probable than reactive trajectories that cross the barrier. The statistical distribution of various trajectories is taken into account in the definition of the transition path ensemble, which assigns the appropriate probability weight to each individual trajectory. Since in transition path sampling one is interested only in transition pathways, i.e., trajectories that connect the stable states, the transition path ensemble excludes trajectories that do not start in A and end in B. The transition path ensemble is then sampled with a Monte Carlo procedure that generates trajectories according to their statistical weight. If the sampling is ergodic, all important pathways will be found and can then be analyzed to yield information on the mechanism but also on the kinetics. In the following sections we will outline the basic principles and algorithms of transition path sampling. For further information on various aspects of transition path sampling we refer the reader to [20–26].

3.2.1. The Transition Path Ensemble

The conceptual starting point of transition path sampling is the definition of the transition path ensemble, a statistical description of all pathways connecting stable states A and B as illustrated in Figure 3-1. Each of these trajectories has the same temporal length t and consists of an ordered sequence of $L = t/\Delta t$ microscopic states separated by a small time step Δt,

$$x(t) \equiv \{x_0, x_{\Delta t}, x_{2\Delta t}, \ldots, x_t\}. \tag{3-1}$$

Such a sequence of states may, for instance, result from a molecular dynamics or Brownian dynamics simulation carried out with a time step Δt. Each microscopic

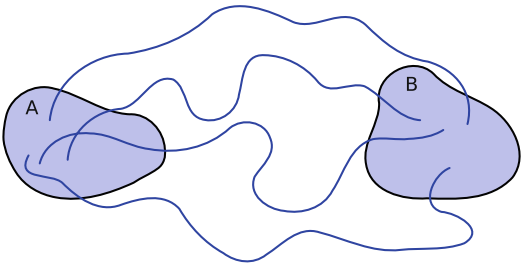


Figure 3-1. The transition path ensemble includes all trajectories of a given length that connect the stables states A and B

state x_τ, or time slice, along a trajectory is a complete copy of the system and, depending on the dynamics considered, includes only the positions or the positions and momenta of all particles. Subsequent time slices on a trajectory are connected by the dynamics of the system. If we denote the short time transition probability from state x_τ at time τ to state $x_{\tau+\Delta t}$ a time step later by $p(x_\tau \to x_{\tau+\Delta t})$, the probability density to observe a given trajectory is given by

$$P[x(t)] = \rho(x_0) \prod_{i=0}^{t/\Delta t-1} p(x_{i\Delta t} \to x_{(i+1)\Delta t}). \tag{3-2}$$

Here, $\rho(x_0)$ is the distribution of the initial conditions from which the trajectories start. Equation (3-2) is valid provided the dynamics of the system is Markovian, i.e., the future time evolution of the system depends only on its current state and not on its past. Most of the kinds of dynamics considered in molecular simulations, including Newtonian dynamics, Langevin dynamics and Monte Carlo dynamics, obey this condition.

The probability density of Eq. (3-2) describes the likelihood of observing a trajectory starting end ending at arbitrary microscopic states. In transition path sampling, however, one is specifically interested only in pathways that are reactive, i.e., that start in A and end in B. This condition on the pathways is included into the statistical description of pathways by multiplying the unrestricted probability density $P[x(t)]$ with the characteristic functions of regions A and B acting on the initial and final time slice of the path, respectively:

$$P_{AB}[x(t)] \equiv h_A(x_0)P[x(t)]h_B(x_t)/Z_{AB}(t). \tag{3-3}$$

The characteristic functions $h_A(x)$ and $h_B(x)$ are defined such that they are unity if their argument is in the respective region and they vanish otherwise. Thus, $h_A(x)$ is given by

$$h_A(x) = \begin{cases} 1 & \text{if} \quad x \in A, \\ 0 & \text{if} \quad x \notin A, \end{cases} \tag{3-4}$$

and $h_B(x)$ is defined analogously. In Eq. (3-3), $P_{AB}[x(t)]$ is normalized by the factor

$$Z_{AB}(t) \equiv \int \mathcal{D}x(t)\, h_A(x_0) P[x(t)] h_B(x_t) \tag{3-5}$$

where the notation

$$\int \mathcal{D}x(t) \equiv \int \cdots \int dx_0 dx_{\Delta t} dx_{2\Delta t} \cdots dx_t, \tag{3-6}$$

familiar from path integrals, indicates an integration over all time slice of the path. The probability density $P_{AB}[x(t)]$, we call it the *transition path ensemble* (TPE), is a statistical description of all pathways of length t that connect the stable states A and B. All pathways that are not reactive are assigned a vanishing weight and thus are not members of the transition path ensemble.

The specific functional form of $P_{AB}[x(t)]$ depends on the distribution of initial conditions, the underlying dynamics and on the definition of the initial and final regions. Depending on the particular situation one considers, the distribution of initial conditions may be the microcanonical or the canonical one. Other distributions are possible as well, including non-equilibrium distributions [27, 28]. The short-time transition probabilities $p(x_\tau \to x_{\tau+\Delta t})$, which enter the espression for the transition path ensemble in Eq. (3-3), depend on the kind of dynamics chosen to model the time evolution of the system. While for a deterministic time evolution such as Newtonian dynamics the transition probabilities are delta functions leading to a highly singular transition path ensemble [29], the transition probabilities are smooth functions for stochastic dynamics, such as the one produced by the Langevin equation [18]. Finally, care must be exercised in the definition of the stable states A and B. These regions, usually defined in configuration space, should be large enough to include all equilibrium fluctuations of the system in the stable states, but should not overlap with their mutual basins of attraction [19].

While the formal definition of the transition path ensemble poses no difficulty, its practical value hinges on ones ability to generate trajectories according to their weight in this ensemble. An efficient way to accomplish exactly that is discussed in the next section.

3.2.2. Monte Carlo in Trajectory Space

In a transition path sampling simulation the transition path ensemble is sampled following the basic idea of a Monte Carlo simulation. The respective procedure is carried out in two basic steps. First, a new trajectory, $x^{(\mathrm{n})}(t)$, is generated from an old one, $x^{(\mathrm{o})}(t)$, for instance using the shooting algorithm described below. Then, the newly generated trajectory is accepted or rejected according to the relative statistical weights of the new and old trajectories. If the new trajectory is accepted, the old trajectory is replaced by the new one. Otherwise, the old one is kept. Iterating

these two basic steps generates a biased random walk in trajectory space, in which trajectories are visited according to their weight in the transition path ensemble.

To ensure that the transition path ensemble is sampled correctly, one requires that detailed balance is obeyed,

$$P_{AB}[x^{(\mathrm{o})}(t)]\pi[x^{(\mathrm{o})}(t) \to x^{(\mathrm{n})}(t)] = P_{AB}[x^{(\mathrm{n})}(t)]\pi[x^{(\mathrm{n})}(t) \to x^{(\mathrm{o})}(t)]. \tag{3-7}$$

Here, $\pi[x^{(\mathrm{o})}(t) \to x^{(\mathrm{n})}(t)]$ is the probability to move from the old path $x^{(\mathrm{o})}(t)$ to the new path $x^{(\mathrm{n})}(t)$ in one Monte Carlo step. This conditions requires that the flow in trajectory space from $x^{(\mathrm{o})}(t)$ to $x^{(\mathrm{n})}(t)$ is exactly compensated by a flow of equal magnitude in the backward direction. If the transition probability π satisfies the detailed balance condition of Eq. (3-7), the Monte Carlo algorithm conserves the transition path distribution $P_{AB}[x(t)]$ and, if ergodic, results in correct sampling of reactive trajectories. For the two-step Monte Carlo procedure described above, the transition probability $\pi[x^{(\mathrm{o})}(t) \to x^{(\mathrm{n})}(t)]$ is given by the product of the probability $P_{\mathrm{gen}}[x^{(\mathrm{o})}(t) \to x^{(\mathrm{n})}(t)]$ to generate the new path from the old one and the probability $P_{\mathrm{acc}}[x^{(\mathrm{o})}(t) \to x^{(\mathrm{n})}(t)]$ to accept the newly generated path,

$$\pi[x^{(\mathrm{o})}(t) \to x^{(\mathrm{n})}(t)] = P_{\mathrm{gen}}[x^{(\mathrm{o})}(t) \to x^{(\mathrm{n})}(t)] \times P_{\mathrm{acc}}[x^{(\mathrm{o})}(t) \to x^{(\mathrm{n})}(t)]. \tag{3-8}$$

Inserting this particular form of the transition probability into the detailed balance condition (3-7) on obtains a condition for the acceptance probability, which can be satisfied using the celebrated Metropolis rule [30], eventually leading to

$$P_{\mathrm{acc}}[x^{(\mathrm{o})}(t) \to x^{(\mathrm{n})}(t)] = h_A[x_0^{(\mathrm{n})}]h_B[x_t^{(\mathrm{n})}] \times \min\left\{1, \frac{P[x^{(\mathrm{n})}(t)]P_{\mathrm{gen}}[x^{(\mathrm{n})}(t) \to x^{(\mathrm{o})}(t)]}{P[x^{(\mathrm{o})}(t)]P_{\mathrm{gen}}[x^{(\mathrm{o})}(t) \to x^{(\mathrm{n})}(t)]}\right\}. \tag{3-9}$$

According to this equation, which provides a general prescription for accepting or rejecting new pathways, a pathway that does not start in A and end in B is immediately rejected. Pathways that are reactive, on the other hand, are accepted with a probability that depends both on the relative weight of the old and the new path in the transition path ensemble as well as on the ratio of the forward and backward generation probabilities.

The specific form of the acceptance probability resulting from Eq. (3-9) depends on the particular way new pathways are generated from old ones. The particular algorithm chosen to do that also controls how rapidly path space is sampled and thus determines the efficiency of the transition path sampling simulation. One path generation method that has proven particularly simple, practical, and efficient is the so-called *shooting algorithm* [29], depicted schematically in Figure 3-2. In this approach, one first randomly selects a time slice $x^{(\mathrm{o})}(t')$, the *shooting point*, from the

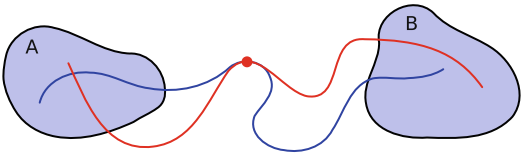


Figure 3-2. In a shooting move, a new trajectory (*blue*) is generated from an old one (*red*) by first randomly choosing a time slice of the old path, and then "shooting off" new trajectory segments forward and backward in time, starting from this shooting point. If the underlying dynamics is deterministic, the shooting point must be modified before the shooting takes place

old path. Then, this shooting point is modified, for instance by adding a random perturbation to the momenta. Starting from the modified shooting point, one integrates the equations of motion of the system forward to time t and backward to time 0 obtaining a complete new trajectory $x^{(\mathrm{n})}(t)$. While for stochastic dynamics the modification step may be omitted, it is strictly necessary for deterministic dynamics. In the latter case, the new trajectory differs from the old one only if the shooting point is modified first. In both cases, the acceptance probability for the shooting move is given by

$$P_{\mathrm{acc}}[x^{(\mathrm{o})}(t) \to x^{(\mathrm{n})}(t)] = h_A[x_0^{(\mathrm{n})}]h_B[x_t^{(\mathrm{n})}] \min\left[1, \frac{\rho(x_{t'}^{(\mathrm{n})})}{\rho(x_{t'}^{(\mathrm{o})})}\right]. \tag{3-10}$$

Thus, non-reactive trajectories are rejected and reactive ones are accepted with a probability that depends only on the equilibrium distribution at the shooting point before and after the modification. The acceptance probability is particularly simple, if the dynamics conserves the energy and the distributions of initial conditions is the microcanonical one:

$$P_{\mathrm{acc}}[x^{(\mathrm{o})}(t) \to x^{(\mathrm{n})}(t)] = h_A[x_0^{(\mathrm{n})}]h_B[x_t^{(\mathrm{n})}]. \tag{3-11}$$

In this case, non-reactive trajectories are rejected and all reactive ones accepted.

For deterministic dynamics, the modification of the shooting point offers the possibility to tune the acceptance probability and, hence, to optimize the efficiency of the simulation. For very small perturbations of the shooting point, the new trajectory retraces the old trajectory to a large degree. Consequently, the new trajectory has a high probability to be reactive and to be accepted. In this regime, most trajectories are accepted, but since subsequent trajectories are very similar, sampling progress is slow. In contrast, very large perturbations of the shooting point lead to new trajectories that markedly differ from the old ones. Nevertheless, the sampling can be inefficient, if most of the new trajectories are non-reactive and are therefore rejected. Optimum sampling efficiency is obtained for shooting point perturbations with a magnitude somewhere between these two extreme cases. This optimum regime is characterized by an average acceptance probability of 20–60% [31].

3.2.3. Analyzing Trajectories

As a result of a transition path sampling simulation, one obtains a collection of reactive pathways that are typical representatives of the transition path ensemble. Extracting information on the transition mechanism from these pathways is, however, often non-trivial. In this section we will review several computational tools that can be used to for this purpose.

A recurrent problem in molecular simulation is to identify those degrees of freedom that capture the essential physics of the process under study and to separate them from the unessential ones that merely act as random noise. For the freezing transition, for instance, it is often unclear whether the size of the crystalline nucleus is sufficient to describe the progress of the transition or if its shape also plays an important role. While for processes occurring in low-dimensional systems with a handful of degrees of freedom, such as a chemical reaction in the gas phase, locating the saddle points on the potential energy surface often yields valuable mechanistic information, the situation is much more involved in high-dimensional systems. Consider, for instance, a chemical reaction in solution. In this case, solvent degrees of freedom may play an important role that is not easily determined from a saddle point analysis. One difficulty is that the number of saddle points grows exponentially with the number of degrees of freedom such that a complete enumeration of the saddle points becomes impractical beyond a certain system size. Perhaps more importantly, the transition of interest is typically not associated with single saddle points that the system must cross on its way from the initial to the final state. In our crystallization example the critical nucleus does not necessarily coincide with any saddle point in the potential energy surface.

By watching the atomic motions during the transitions with a molecular viewing program on a computer, one may gain some information about the process of interest. While it is often useful and stimulationg to do so, important details of the mechanism, which can be best captured in form of a reaction coordinate, may remain hidden to the eye. A reaction coordinate $q(r)$ is a function of the configuration r of the system, which quantifies the progress of the reaction. In the case chemical reactions, for instance, bond angles or bond lengths may serve as a reaction coordinate; for a folding protein, the number of native contacts may provide a measure for the folding progress. A good reaction coordinate should tell us how far the reaction has proceeded and what is likely to happen next. The concept of the quality of a reaction coordinate can be made more precise by considering the so called *committor* [32, 33], introduced by Onsager as splitting probability [34] and known as p_{fold} in the context of protein folding [35]. The committor $p_B(r)$, which can be defined in configuration space or in phase space [12], is the probability that a trajectory starting from r reaches B rather than A first. As indicated in Figure 3-3, the committor can be calculated for a particular configuration r by initiating a number of short trajectories from that configuration and determining the fraction of trajectories that end up in B rather than A. A committor value close to *1* indicates that trajectories started from r are very likely to relax into B. While they do not necessarily lie in B itself, such configurations are strongly committed to B and can be viewed to be

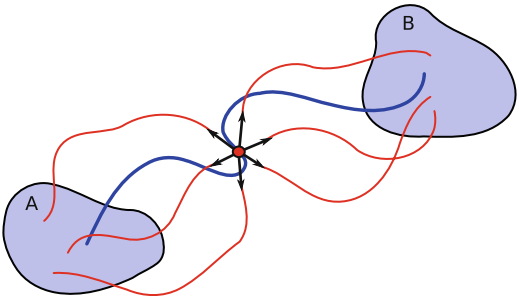


Figure 3-3. To estimate the committor p_B for a particular configuration r (*red point*), one starts n trajectories from r with random initial momenta and determines the number n_B of trajectories that reach B rather then A. The committor is then given by the fraction $p_B = n_B/n$.

part of the basin of attraction of region B. Committor values close to 0, on the other hand, characterize configurations that will most likely relax into region A.

Configurations with $p_B(r) = p_A(r) = 1/2$ play a special role because they can be identified as transition states, from which both stable states are accessible with equal probability [36–40]. This statistical concept of a transition state, which generalizes the conventional definition of a transition state as saddle point on the potential energy surface, is applicable also to the complex high-dimensional systems of interest here. By determining all configurations with $p_B = 1/2$ on transition pathways one obtains the so-called *transition state ensemble* (TSE). Comparison of configurations belonging to the transition states ensemble with those from the stable states can yield important information on the transition mechanism.

As asserted above, a good reaction coordinate should provide a measure for the progress of a particular reaction. In this sense, the committor is the perfect reaction coordinate as it exactly specifies how far the reaction has proceeded and what is likely to happen next [41]. Unfortunately, the committor is very unspecific and does not automatically lead to insight into the mechanism in terms of physical variables. Furthermore, the committor is not easy to evaluate numerically such that it is impractical to use the committor, for instance, as reaction coordinate in a transition state theory calculation. However, the committor can be used as a criterion for distinguishing between good and poor reaction coordinates. If $q(r)$ is a good reaction coordinate, its value determines the progress of the reaction and the committor is completely determined by $q(r)$:

$$p_B(r) = p_B[q(r)]. \tag{3-12}$$

For a poor reaction coordinate, on the other hand, the value of the reaction coordinate does not predict the committor and the above relation does not hold.

The fact that a good reaction coordinate determines the committor can be used to test a proposed reaction coordinate $q(r)$. One may, for instance, expect that for the freezing transition of a particular material the size of the crystalline nucleus is a good reaction coordinate. To examine the quality of this coordinate, one first

computes the free energy F as a function of q. If $q(r)$ has any relation to the rare event of interest, $F(q)$ is expected to be bistable with a barrier at $q = q^*$ separating the free energy minima corresponding to the stable states A and B. Since the value of the reaction coordinate completely specifies the committor, all configurations with the same reaction coordinate should also have the same committor. So if one generates configurations for a fixed value of $q(r)$, for instance with a constrained molecular dynamics simulation, and computes the committor value for each of this configurations, the resulting distribution of committor values, $P(p_B)$, should be delta peaked at $p_B(q)$. In particular, configurations with reaction coordinate $q(r) = q^*$ corresponding to the barrier top should all have the committor $p_B(q^*)$ and hence the corresponding committor distribution has a sharp peak at $p_B(q^*)$. For a good reaction coordinate, the barrier top coincides with the transition state ensemble and the peak is located at $p_B = 1/2$. In contrast, a poor reaction coordinate does not determine the value of the committor and hence does not lead to sharply peaked committor distributions. Rather, the committor distribution calculated for configurations constrained to the barrier top typically is bimodal with peaks at 0 and 1. Thus, any committor distribution without a single sharp peak is an indication of an inadequate reaction coordinate. An analysis based on committor distributions has been used to reveal the reaction coordinate of ionic dissociation [32], biomolecular isomerization [33], and the freezing transition [42].

3.2.4. Calculating Rate Constants

Reaction rate constants, describing the kinetics of processes involving rare events, are often measured empirically and thus provide an important way to establish close contact between molecular simulation and experiment. Since pathways harvested in a transition path sampling simulation are truly dynamical trajectories, they can be used to compute such reaction rate constants. The transition path ensemble, however, is restricted only to short trajectory segments during which the transition of interest occurs such that reaction rates cannot be directly extracted from these pathways. While the relative probabilities of different reactive trajectories are correctly described by the transition path ensemble, the information on the probability of observing a reactive event at all (as opposed to no event) is not contained in this path ensemble. Therefore, the fundamental problem in calculating reaction rate constants with transition path sampling consists in estimating the relative weight of the reactive trajectories with respect to all possible trajectories. Several approaches to do that have been suggested in the past and we will briefly survey them in this section. A more detailed review of these methods is provided in [13].

The link between the microscopic dynamics of the system and its phenomenological description in terms of reaction rate constants is provided by the time correlation function

$$C(t) \equiv \frac{\langle h_A(x_0) h_B(x_t) \rangle}{\langle h_A \rangle}. \tag{3-13}$$

Here, the angular brackets $\langle \cdots \rangle$ denote an equilibrium average. This time correlation function equals the conditional probability to observe the system in region B at time t provided it was in A at time 0. In the case of two-state kinetics, $C(t)$ approaches its asymptotic value exponentially,

$$C(t) = \langle h_B \rangle (1 - \exp(-t/\tau_{\text{rxn}})). \tag{3-14}$$

where the relaxation time τ_{rxn} is related to the forward and backward reaction rate constants k_{AB} and k_{BA} by

$$\tau_{\text{rxn}}^{-1} = k_{AB} + k_{BA}. \tag{3-15}$$

The exponential behavior of Eq. (14) cannot be valid for very short times. If regions A and B are not adjacent and there is a gap between them, the system will need a minimum time τ_{mol} to cross this gap. Only for times larger than τ_{mol} can exponential behavior set in as expected from the solution of the phenomenological rate equations. If there is a separation of time scales, i.e., if there is a time regime such that $\tau_{\text{mol}} < t \ll \tau_{\text{rxn}}$, the exponential growth can be approximated by a linear behavior such that

$$C(t) \approx k_{AB} t. \tag{3-16}$$

Equivalently, the time derivative $k(t) = dC(t)/dt$ reaches a plateau for times $\tau_{\text{mol}} < t \ll \tau_{\text{rxn}}$ [4]. Thus, knowledge of the time correlation function $C(t)$ is sufficient for a calculation of the forward reaction rate constant k_{AB}.

One transition path sampling approach for the computation of reaction rate constants consists in determining the time correlation function $C(t)$ using free energy calculation techniques [31, 25]. In this method, one rewrites $C(t)$ as

$$C(t) = \frac{\int \mathcal{D}x(t) P[x(t)] h_A(x_0) h_B(x_t)}{\int \mathcal{D}x(t) P[x(t)] h_A(x_0)}, \tag{3-17}$$

and observes that $C(t)$ can be viewed as a ratio of two partition functions. In Eq. (3-17), the numerator is the partition function of all pathways starting in A at time 0 and ending in B at time t. The denominator, on the other hand, is the partition function of all pathways starting in A without and restriction on where they end. Hence, the ratio of these partition function is related to the "reversible work" $W_{AB}(t)$ required to transform the ensemble of trajectories with free endpoints into that with endpoints in B, $C(t) = \exp[-W_{AB}(t)]$. The reversible work $W_{AB}(t)$, a free energy in trajectory space, can be calculated with standard free energy methods such as umbrella sampling [31], thermodynamic integration [43], or even Jarzynski fast switching [44]. In these calculations, one starts with a final region that encompasses the entire configuration space and then successively shrinks it to the desired size. Since the calculation of the reaction rate constant requires the calculation of the time derivative of $C(t)$, in principle several of these path free energy calculations have to

be carried out for different path lengths t. This costly operation can be avoided, by calculating $C(t)$ in two steps. First, the time correlation function is calculated with a free energy procedure for one particular time t'. In the second step, the path free energy required to change the path length from t' to t is calculated. This can be done for all values of t up to a maximum time $t_{\max}$ in one single regular transition path sampling simulation [31]. Combining the results of these two calculations one obtains the correlation function $C(t)$ from 0 to $t_{\max}$ and the reaction rate constant can then be extracted from it.

An alternative transition path sampling algorithm for the calculation of reaction rate constants was proposed by Bolhuis and collaborators and named transition interface sampling (TIS) [45, 46]. In this method, pathways of variable length are used which leads to a reduced numerical effort with respect to the method described above. Transition interface sampling, however, is based on an additional assumption about correlated transitions between the stable states. The method rests on the concept of the "overall states" $\mathcal{A}$ and $\mathcal{B}$. Overall state $\mathcal{A}$ consists of points in A plus all points that originate from A in the sense that a trajectory going through such points reaches A before B if followed backwards in time. (This definition is valid only for deterministic trajectories.) Overall state $\mathcal{B}$ is defined analogously. The two overall states $\mathcal{A}$ and $\mathcal{B}$ cover the entire phase space with a possibly very complicated boundary separating them. If one now considers the time correlation function

$$\mathcal{C}(t) \equiv \frac{\langle h_{\mathcal{A}}(x_0) h_{\mathcal{B}}(x_t)\rangle}{\langle h_{\mathcal{A}}\rangle} \tag{3-18}$$

and evaluates the corresponding time derivative in the transition state theory approximation (recrossings are excluded [20]), one obtains the expression

$$k_{AB} = \frac{\langle \phi_{AB}\rangle}{\langle h_{\mathcal{A}}\rangle} \tag{3-19}$$

for the rate constant. Here, $\langle \phi_{AB}\rangle$ is the effective positive flux into region B, i. e., the average flux into B due to trajectories coming directly from A. Thus, for a trajectory connecting A with B only the first entry of the trajectory into B contributes to the effective positive flux. Since the above expression of the rate constant was obtained from a TST-approximation for the overall states $\mathcal{A}$ and $\mathcal{B}$, the underlying approximation is that there are no correlated transitions from A to B and back, a condition that is often but not always satisfied. For stables states A and B defined in configuration space as is customary, chemical reactions occurring in the energy diffusion regime, for instance, may violate this assumption.

In principle, the effective positive flux $\langle \phi_{AB}\rangle$ could be calculated from a long molecular dynamics trajectory by counting the number of first entries into B occurring per time unit. Of course, rare events make this direct approach impractical. To calculate the effective positive flux, Bolhuis and collaborator have therefore developed a technique based on a sequence of non-intersecting interfaces that span the region between A and B [45, 46]. The spacing between these interfaces is selected

such that a trajectory crossing interface i coming from A has a non-vanishing probability of also crossing the interface $i + 1$. The effective positive flux can then be expressed as the product of the average positive flux through the surface of A with the product of all these crossing probabilities. The effective positive flux is thus given by the average flux out of A multiplied with the probability of these exit trajectories to eventually reach B. This probability can be calculated from transition path sampling simulations carried out separately for each interface. The ensemble sampled in these simulations consists of trajectories with varying length starting in A, reaching interface i, and then going back to A or on to cross interface $i + 1$. To date, transition interface sampling has been used to calculate reaction rate constants for the freezing transition in simple liquids [42] and several biomolecular isomerizations [47, 48].

For very long and diffusive barrier crossing processes the efficiency of transition interface sampling simulations can be considerably increased by exploiting the loss of correlations along individual pathways. This idea is used in the partial path transition interface sampling (PPTIS) method [49]. Another method similar in spirit to the transition interface sampling algorithm is the so-called forward flux sampling method (FFS), which can be applied also to non-equilibrium systems in which the stationary phase space distribution is unknown [50–52].

3.3. A TPS ALGORITHM FOR NANOCRYSTALS IN A PRESSURE BATH

In the previous sections we have given a brief outline of the main concepts and algorithms of transition path sampling. In this and the following sections we will explain how to apply these techniques for the simulation of pressure induced structural phase transitions in semiconductor nanocrystals.

3.3.1. Ideal Gas Pressure Bath

A central part of a computer simulation of a nanoparticle under pressure is the pressure bath. In experimental studies, ethylcyclohexane [53–55] and ethyl-pyridine [14] have been used as a pressure medium and solvent for CdSe nanocrystals. These substances guarantee pressurization of the crystals under quasi-hydrostatic conditions up to pressures of 10 GPa [14]. In a computer simulation of a single nanocrystal under pressure, the number of particles in the pressure bath exceeds the number of crystal atoms by far. Apart from being able to apply hydrostatic pressure, a simulated pressure medium therefore also should be efficient in terms of computer time.

In the first computational study of a pressure-induced structural transformation of a nanocrystal, Martoňák, Molteni and Parrinello used a liquid of soft spheres to transmit hydrostatic pressure on a small silicon cluster [56–59]. In their scheme, the pressure is controlled by tuning the parameter of the interaction of particles in the pressure medium. This is possible because the equation of state of the soft sphere

system is known from computer simulation [60]. In this method, fluctuations of the volume of the crystal, in particular the large volume change accompanying the structural transformation, affect the pressure. For precise control over the pressure, the volume of the nanocrystal therefore must be known.

In a different approach, Morgan and Madden studied structural transformations of ionic nanocrystals using particles of the Lennard-Jones type as a pressure bath [61–64]. The system was coupled to the Andersen baro- and thermostat [65, 66], which modifies the dynamics of the system to achieve conditions of constant pressure an temperature. A similar approach was taken by Vashishta and coworkers in their study of GaAs nanocrystals [67, 68] and CdSe nanorods [69].

Finally, methods have been proposed that do not make use of an explicit pressure medium, but use a barostat that changes the equations of motion of crystal atoms by coupling to the volume of the crystal. Although computationally cheap, these methods do not directly model the experimental situation and results can depend strongly on the specific definition of the cluster volume [70].

In the following, we concentrate on the ideal gas barostat, a method developed in our group [15, 16]. In this scheme, a nanocrystal is surrounded by a cloud of non-interacting, hence ideal, particles, that transmit the pressure on the crystal via a simple, purely repulsive interaction potential. No fixed system volume is used, nor is the number of gas particles a constant. Instead, the shape and size of the ideal gas atmosphere is dynamically adjusted to adapt to shape and volume changes of the crystal. To achieve the correct statistics of an ideal gas at the desired pressure and temperature, the proper flow of gas particles through the boundary of the atmosphere is modeled stochastically. In effect, the method allows the simulation of an arbitrarily shaped, microscopic subvolume of a macroscopic ideal gas (see Figure 3-4 for an illustration of the method).

3.3.1.1. Algorithm

For an implementation of the ideal gas barostat [15, 16], the simulation box is divided in cubic cells of side length $l_{\text{cell}} \geq r_{\text{cut}}$. Here, r_{cut} is the cutoff distance of the interaction between gas particles and crystal atoms,

$$u(r) = \begin{cases} \varepsilon \left[(r/\sigma)^{-12} - (r_{\text{cut}}/\sigma)^{-12} \right] & \text{if } r < r_{\text{cut}} , \\ 0 & \text{if } r \geq r_{\text{cut}} . \end{cases} \tag{3-20}$$

The parameters ε and σ should be chosen large enough to prevent gas particles from penetrating the crystal; a convenient value for the cutoff distance is $r_{\text{cut}} = 2\sigma$. At any given time, the volume occupied by ideal gas particles consists of those cells that hold possible interaction partners of crystal atoms. (If a given cell is occupied by a crystal atom, the cell itself and all 26 neighbor cells are considered as a part of the atmosphere.)

If a gas particle leaves the so-defined atmosphere, it is removed from the simulation. The resulting outward flow of gas particles is balanced by new gas particles

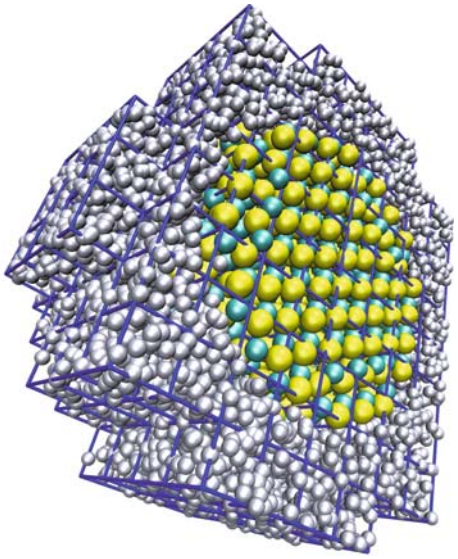


Figure 3-4. Cross section of a CdSe nanocrystal (*cyan and yellow*) immersed in the pressure bath of ideal gas particles (*gray*). The gas occupies only a thin layer, or atmosphere, around the nanocrystal. Built from rectangular cells (*blue grid*), this atmosphere adapts to the shape of the crystal and thus always provides the amount of gas particles necessary to exert the desired pressure

that are introduced at the boundary of the atmosphere. To ensure correct statistics, on average

$$\bar{n} = (2\pi m k_\mathrm{B} T)^{-1/2} P l_\mathrm{cell}^2 \Delta t \tag{3-21}$$

particles need to enter the atmosphere through any given side face of a cell during a time interval of length Δt. Here, P is the pressure, T is the temperature, and k_B is the Boltzmann constant. While two of the three velocity components of these particles follow a Maxwell-Boltzmann distribution, the velocity component $v_\perp$ perpendicular to the side face through which the particles enter is distributed according to

$$p(v_\perp) = \frac{m}{k_\mathrm{B} T} v_\perp \exp\left(-\frac{m v_\perp^2}{2 k_\mathrm{B} T}\right) . \tag{3-22}$$

When the crystal changes its shape or position in the simulation box, parts of the atmosphere cease to be nearest neighbors of crystal-filled cells. As these parts, or cells, are no longer needed to apply pressure on the crystal, they are removed from the simulation. On the other hand it is necessary to add new gas-filled cells to the atmosphere when crystal atoms enter cells not occupied before. On average,

$$\bar{N} = P l_\mathrm{cell}^3 / k_\mathrm{B} T \tag{3-23}$$

gas particles should be put into newly added cells, with positions equally distributed over the cell volume.

The following molecular dynamics scheme propagates the crystal and gas particles for one time step of length Δt. The integration of the equations of motion is carried out with the Velocity Verlet algorithm [71].

MD algorithm

1. Propagate the positions of crystal atoms for one time step (first part of the Verlet integrator), using forces calculated in the last time step.
2. Check if the gas atmosphere needs to be updated by removing or adding gas-filled cells.
3. Remove cells (and gas particles therein) from the atmosphere, if they are no longer needed.
4. Propagate the positions of all gas atoms for one time step (first part of the Verlet integrator), using forces calculated in the last time step. Remove particles that leave the atmosphere.
5. Insert new gas particles on the boundary of the atmosphere. The number of particles placed on a given side face of a cell is Poisson-distributed with average $\bar{n}$. Positions of added particles are uniformly distributed. Velocity components are distributed according to $p(v_\perp)$ for the direction perpendicular to the side face, and follow a Maxwell-Boltzmann distribution for the in-plane directions. Propagate the newly added particles for a time interval uniformly distributed between zero and Δt. Remove particles that leave the atmosphere in this step.
6. If necessary, add new cells to the atmosphere and fill them with a Poisson-distributed number of gas particles with average $\bar{N}$. Positions are uniformly distributed and velocities are drawn from a Maxwell-Boltzmann distribution.
7. Compute the forces.
8. Using the newly computed forces, propagate the velocities of all particles, gas and crystal, for one time step (second part of the Verlet integrator).

It can be shown analytically that this algorithm satisfies detailed balance and therefore conserves the equilibrium distribution of a system at constant pressure and temperature [16].

3.3.2. Simple Shooting Moves

The ideal gas barostat discussed in the last section can be easily incorporated into the framework of a TPS simulation [16]. As discussed in Section 3.2.2, the acceptance criterion for shooting moves takes a particularly simple form if the underlying dynamics of the system satisfy detailed balance. This is the case for the ideal gas

barostat. Furthermore, there is no need to actively modify the phase space vector of the system at the shooting point, because the stochastic element of the barostat – new gas particles are introduced to the simulation in a random fashion – will lead to divergent trajectories typically within a few thousand time steps. Also, as the dynamics of the system is not purely deterministic, one-way shooting moves, where only one part of the trajectory is resampled, may be used to increase the acceptance of newly generated trajectories. The following shooting algorithm therefore correctly samples the transition path ensemble of a nanoparticle in a pressure bath.

Shooting move

1. Choose a shooting point along the given trajectory.
2. Choose a direction of time in which to propagate the system from the shooting point.
3. From the chosen shooting point, integrate the equations of motions, either forward or backward in time, until a complete new trajectory is obtained.
4. Accept the trajectory, if it still starts in stable state A and ends in state B; reject otherwise.

To enhance the efficiency of shooting moves, shifting moves should be applied in the customary way [19].

3.4. THE WURTZITE TO ROCKSALT TRANSFORMATION IN CDSE NANOCRYSTALS

In this section, we discuss the application of TPS to the case of a structural transformation in CdSe nanocrystals. The bulk semiconductor CdSe transforms from the four-coordinated, hexagonal wurtzite structure to the six-coordinated, cubic rocksalt structure at pressures around 2.5 GPa [72]. The analogous transformation in CdSe nanocrystals in the size range of a few nanometers has become the paradigmatic example for size effects in phase transitions. In a series of experiments, Alivisatos and coworkers were able to clarify much of the kinetics and thermodynamics of these transitions. The thermodynamic transition pressure increases with decreasing crystal size, in agreement with the notion of different surface free-energies of the two crystal structures [73, 14]. The kinetic barriers to the transition, characterized by activation energies and volumes, also show strong dependence on crystal size and pressure, indicating that transformations proceed through single nucleation events and subsequent growth [74, 53, 55, 75].

Although it is possible to exclude certain transition routes from experimental findings [54], the time and space resolution of experiments is insufficient for an understanding of the transformation mechanism on the atomistic scale. Only through detailed knowledge of atomistic transition pathways, however, will it

be possible to meet promising technological challenges like the stabilization of metastable structures through control of crystal shape and surface configuration. In the following, we discuss the use of molecular dynamics computer simulation, and in particular transition path sampling, in identifying the transformation mechanisms of CdSe nanocrystals.

3.4.1. Straightforward MD Simulations

A typical molecular dynamics simulation of a nanocrystal under pressure proceeds in the following way, which is similar to the experimental procedure. A single crystal is constructed from a bulk lattice and thermalized at ambient conditions in the pressure bath. In small steps, pressure is increased and the crystal is allowed to relax at every pressure level typically for a few picoseconds. A structural transformation of the crystal can be conveniently identified by visual inspection, or by monitoring quantities like the potential energy or coordination number of crystal atoms. Using the ideal gas barostat and the approach outlined above, we have studied the wurtzite to rocksalt transformation in CdSe nanocrystals consisting of 100–5000 atoms [76]. The crystals were built according to electron microscopy images of experimentally used crystals [77, 78] and had two different shapes, which markedly influenced the transformation behavior. While spherical crystals with disordered, strongly reconstructed surfaces transformed directly from wurtzite to rocksalt, faceted crystals with well-defined, stable surfaces transformed in two steps: At intermediate pressures, a compression of the crystal along the hexagonal c-axis produces the five-coordinated *h*-MgO structure. This metastable intermediate is unstable in bulk CdSe [79–81] and stabilized in faceted crystals by a favorable surface free energy [76, 69, 82] (see Figure 3-5).

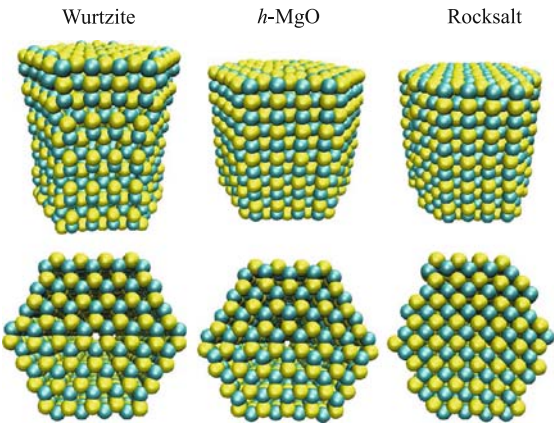


Figure 3-5. A faceted $Cd_{528}Se_{528}$ nanocrystal in the wurtzite, *h*-MgO, and rocksalt structures. The *crystals* are shown both from the *side* and the *top,* looking down the hexagonal *c*-axis. Note that the intermediate *h*-MgO structure is obtained from the wurtzite configuration by a simple flattening of the puckered (001) layers, which leaves the hexagonal structure unchanged otherwise

The transformation from wurtzite to rocksalt in spherical crystals and from *h*-MgO to rocksalt in faceted follow similar transition routes. Two main mechanisms could be identified: While in most cases the transformation proceeds through the sliding of parallel (100) crystal planes, a few crystals transformed through a consecutive flattening, or compression, of (100) layers in the [100] direction. The latter mechanism is similar to a transition pathway previously proposed by Alivisatos and coworkers [14].

Straightforward MD simulations like the one discussed here take place on timescales many orders of magnitude smaller than the duration of typical experiments. As a result, pressures of up to 10 GPa have to be used to induce the transformation (depending on crystal size and shape), while transformation pressures of about 5 GPa are reported from experiments [53]. At elevated pressures, the kinetic barrier of the transition essentially vanishes and the transformation becomes observable on the picosecond timescale of the simulation. Under these extreme conditions, transformations proceed in violent ways: simultaneous nucleation from different sites on the crystal and the formation of multiple grains are frequently observed [61, 68, 76] – a direct comparison with experiments becomes increasingly difficult.

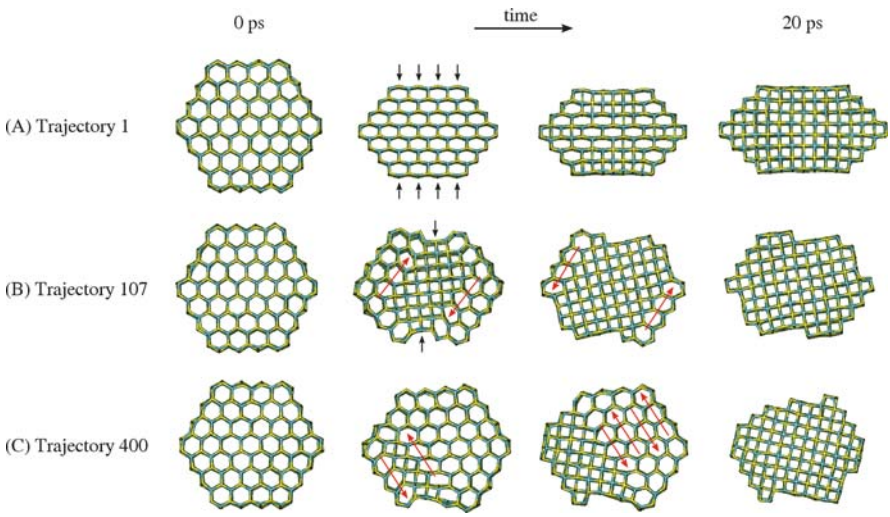


Figure 3-6. Snapshots along three trajectories from a TPS simulation of a $Cd_{528}Se_{528}$ nanocrystal at 3 GPa. The crystal is viewed along its hexagonal *c*-axis; the pressure bath is not shown. In each trajectory, time passes from *left* to *right.* The crystal starts in the *h*-MgO structure and, at the end of the trajectory, has transformed to the rocksalt structure. (**a**) In the first (artificially created) trajectory, the crystal is transformed via a collective motion of all crystal atoms and is compressed in one of three equivalent [100] directions, indicated by *black arrows*. (**b**) After 107 iterations of the algorithm, the mechanism has changed. Starting in a fashion similar to the initial trajectory, the final steps of the transformation occur through sliding of (100) planes, as indicated by *red arrows.* (**c**) In pathway 400, trace of the initial mechanism disappears and the entire transformation, starting on the crystal surface, proceeds via the sliding-planes mechanism. Reprinted with permission from J. Chem. Phys. 127, 154718 (2007). Copyright 2007, American Institute of Physics

3.4.2. TPS Reveals the Main Mechanism

In a TPS simulation, these problems are avoided and quantitative information about the relative probability of different mechanisms can be obtained under conditions close to experiments. To clarify the role of the two main mechanisms observed with straightforward MD, we studied the transformation from *h*-MgO to rocksalt in faceted CdSe nanocrystals with the TPS-algorithm discussed in Section 3.3 at pressures between 2 and 3.5 GPa [16]. As a first trajectory, a necessary ingredient to start a TPS simulation, we constructed a pathway that resembles the compression mechanism discussed by Alivisatos and coworkers [14] and illustrated in Figure 3-6a. This specific mechanism, although observed in a few of our high-pressure MD simulations, is highly unfavored at lower pressures: within a thousand iterations of the shooting algorithm, the compression mechanism is lost completely and the simulation converges towards the sliding-planes mechanism illustrated in Figure 3-6c. When continued, the simulation explores equivalent pathways, where the transformation starts from different sites on the crystal surface or proceeds along different sets of parallel (100) planes. The characteristics of the sliding-planes mechanism, however, remain the same. In particular, return to the initial compression mechanism is never observed. This result strongly indicates that the sliding-planes mechanism is the favored transformation pathway for faceted CdSe nanocrystals.

3.5. CONCLUDING REMARKS

Conducting a computer simulation under near-experimental conditions (for instance, using transition path sampling) is a prerequisite but not a guarantee for observing transformation pathways relevant in the real system. For nanocrystals under pressure, the role of different surface passivation agents, defects in the crystal lattice, or the dynamics and composition of the pressure bath is difficult to assess experimentally and even harder to model realistically in a simulation. The relevance of mechanisms observed in a computer simulation can therefore only be established by comparison with experimental data. Transition path sampling offers various techniques to do so. First, calculation of the rate constant allows direct contact with experiments. Though computationally rather expensive, the calculation of rate constants in the framework of transition path sampling (as discussed in Section 3.2.4) does not need a priori knowledge of the reaction coordinate and thus is free of any bias. One drawback that makes a direct comparison with experiments via the rate constant difficult is the fact that rate constants often strongly depend on the quality of empirical potentials. With the development of ever more sophisticated pair potentials and the increasing range of problems that can be tackled ab initio, we nevertheless expect this approach to become fruitful for many systems of interest. A second way to establish contact with empirical data is offered by committor analysis and the determination of the transition state ensemble (as discussed in Section 3.2.3). As the central part of a rare event, the transition state not only offers direct insight into the relevant degrees of freedom governing the transformation, but can

also be quantified in terms of activation energy and activation volume, quantities that are readily accessible in experiments.

In summary, transition path sampling is a versatile and efficient set of computational techniques for the study of rare events in complex systems. It has has been successfully applied to a broad range of problems from material science to molecular biology; it can be adapted to clarify the transformation details of a simple chemical reactions as well as solid-solid phase transitions, occurring in bulk and nanoscale materials. As computers advance, the range and complexity of systems to which transition path sampling can be fruitfully applied increases and a more realistic modelling of the experimental situation becomes possible in many cases.

ACKNOWLEDGMENTS

This work was supported by the Austrian Science Fund (FWF) under Grant No. P20942-N16 and within the Science College "Computational Materials Science" under grant W004.

REFERENCES

1. G. M. Torrie and J. P. Valleau, J. Comp. Phys. **23**, 187 (1977).
2. J. Kirkwood, J. Chem. Phys. **3**, 300 (1935); E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, Chem. Phys. Lett. **156**, 472 (1989).
3. C. H. Bennett, in Algorithms for Chemical Computations, ACS Symposium Series No. **46**, edited by R. Christofferson, American Chemical Society, Washington, D.C. (1977).
4. D. Chandler, J. Chem. Phys. **68**, 2959 (1978).
5. H. Eyring, J. Chem. Phys. **3**, 107 (1935).
6. E. Wigner, Trans. Faraday Soc. **34**, 29 (1938).
7. A. Laio and M. Parrinello, Proc Natl. Acad. Sci. USA **99**, 12562 (2002).
8. A. F. Voter, F. Montalenti, and T. C. Germann, Annu. Rev. Mater. Res. **32**, 321 (2002).
9. G. Hummer and I. G. Kevrekidis, J. Chem. Phys. **118**, 10762 (2003).
10. H. Jonssón, G. Mills, and K. W. Jacobsen, in Classical and Quantum Dynamics in Condensed Phase Simulations, edited by B. J. Berne, G. Ciccotti, and D. F. Coker, p. 385, World Scientific, Singapore (1998).
11. W. E, W. Ren, and E. Vanden-Eijnden, Phys. Rev. B **66**, 052301 (2002).
12. W. E, W. Ren, and E. Vanden-Eijnden, Chem. Phys. Lett. **413**, 242 (2005).
13. C. Dellago and P. G. Bolhuis, Adv. Poly. Sci. **221**, 167 (2008).
14. S. H. Tolbert and A. P. Alivisatos, J. Chem. Phys. **102**(11), 4642 (1995)
15. M. Grünwald and C. Dellago, Mol. Phys. **104**, 3709 (2006)
16. M. Grünwald P. L. Geissler, and C. Dellago, J. Chem. Phys. **127**, 154718 (2007)
17. M. Grünwald and C. Dellago. Nano Letters **9**, 2099 (2009).
18. C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**(5), 1964 (1998).
19. C. Dellago, P. G. Bolhuis, and P.L. Geissler, Adv. Chem. Phys. **123**, 1 (2002).
20. C. Dellago, P. G. Bolhuis, and P. L. Geissler, in Computer Simulations in Condensed Matter: from Materials to Chemical Biology, edited by M. Ferrario, G. Ciccotti, and K. Binder, p. 349, Springer Lecture Notes in Physics, New York (2006).
21. P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Ann. Rev. Phys. Chem. **53** 291 (2002).

22. C. Dellago, P. G. Bolhuis, and P. L. Geissler, Adv. Chem. Phys. **123** 1 (2002).
23. C. Dellago and D. Chandler, in Molecular Simulation for the Next Decade, ed. by P. Nielaba, M. Mareschal, and G. Ciccotti, p. 321, Springer, Berlin (2002).
24. C. Dellago, in Handbook of Materials Modeling, edited by S. Yip, p. 1585, Springer, Berlin (2005).
25. C. Dellago, in Free Energy Calculations: Theory and Applications in Chemistry and Biology, edited by A. Pohorille and C. Chipot, Springer, Berlin (2007).
26. C. Dellago and P. G. Bolhuis, Top. Curr. Chem. **268**, 291, edited by M. Reiher, Springer (2007).
27. G. E. Crooks and D. Chandler, Phys. Rev. E **64**, 026109 (2001).
28. P. L. Geissler and D. Chandler, J. Chem. Phys. **113**, 9759 (2000).
29. P. G. Bolhuis, C. Dellago, and D. Chandler, Faraday Discuss. **110**, 421 (1998).
30. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
31. C. Dellago, P. G. Bolhuis, and D. Chandler, J. Chem. Phys. **110**, 6617 (1999).
32. P. L. Geissler, C. Dellago, and D. Chandler, J. Phys. Chem. B **103**, 3706 (1999).
33. P. G. Bolhuis, C. Dellago, and D. Chandler, Proc. Natl. Acad. Sci. USA **97**, 5877 (2000).
34. L. Onsager, Phys. Rev. **54**, 554 (1938).
35. R. Du, V. Pande, A. Y. Grosberg, T. Tanaka, and E. I. Shakhnovich, J. Chem. Phys. **108**, 334 (1998).
36. D. Ryter, Physica A **142**, 103 (1987).
37. A. Berezhovski and A. Szabo, J. Chem. Phys. **125**, 104902 (2006).
38. M. M. Klosek, B. J. Matkowsky, and Z. Schuss, Ber. Bunsenges. Phys. Chem. **95**, 331 (1991).
39. E. Pollak, A. M. Berezhkovskii, and Z. Schuss, J. Chem. Phys. **100**, 334 (1994).
40. P. Talkner, Chem. Phys. **180**, 199 (1994).
41. G. Hummer, J. Chem. Phys. **120**, 516 (2004).
42. D. Moroni, P. R. ten Wolde, and P. G. Bolhuis, Phys. Rev. Lett. **94**, 235703 (2005).
43. C. Dellago and P. G. Bolhuis, Mol. Sim. **30**, 795 (2004).
44. P. L. Geissler and C. Dellago, J. Phys. Chem. B **108**, 6667 (2004).
45. T. S. van Erp, D. Moroni, and P. G. Bolhuis, J. Chem. Phys. **118**, 7762 (2003).
46. T. S. van Erp and P. G. Bolhuis, J. Comp. Phys. **205**, 157 (2005).
47. P. G. Bolhuis, P Natl Acad Sci USA **100**, 12129 (2003).
48. P. G. Bolhuis, Biophys J **88**, 50 (2005).
49. D. Moroni, P. G. Bolhuis, and T. S. van Erp, J. Chem. Phys. **120**, 4055 (2004).
50. R. J. Allen, P. B. Warren, and P. R. ten Wolde, Phys. Rev. Lett. **94**, 018104 (2005).
51. R. J. Allen, D. Frenkel, and P. R. ten Wolde, J. Chem. Phys. **124**, 024102 (2006).
52. R. J. Allen, D. Frenkel, and P. R. ten Wolde, J. Chem. Phys. **124**, 194111 (2006).
53. K. Jacobs, D. Zaziski, E. C. Scher, A. B. Herhold, and A. P. Alivisatos, Science **293**(5536), 1803 (2001)
54. J. N. Wickham, A. B. Herhold, and A. P. Alivisatos, Phys. Rev. Lett. **84**(5), 923 (2000)
55. K. Jacobs, J. Wickham, and A. P. Alivisatos, J. Phys. Chem. B **106**(15), 3759 (2002)
56. R. Martoňák, C. Molteni, and M. Parrinello, Phys. Rev. Lett. **84**(4), 682 (2000)
57. R. Martoňák, C. Molteni, and M. Parrinello, Comput. Mater. Sci. **20**, 293 (2001)
58. C. Molteni, R. Martoňák, and M. Parrinello, J. Chem. Phys. **114**(12), 5358 (2001)
59. R. Martoňák, L. Colombo, C. Molteni, and M. Parrinello, J. Chem. Phys. **117**(24), 11329 (2002)
60. W. G. Hoover, M. Ross, K. W. Johnson, D. Henderson, J. A. Barker, and B. C. Brown, J. Chem. Phys. **52**(10), 4931 (1970)
61. B. J. Morgan and P. A. Madden, Nano Lett. **4**(9), 1581 (2004)
62. B. J. Morgan and P. A. Madden, Phys. Chem. Chem. Phys. **8**, 3304 (2006)
63. B. J. Morgan and P. A. Madden, J. Phys. Chem. C **111**(18), 6724 (2007)
64. B. J. Morgan, Phys. Rev. B **78**, 024110 (2008)
65. H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980)

66. G. J. Martyna, D. J. Tobias, and M. L. Klein, J. Chem. Phys. **101**(5), 4177 (1994)
67. S. Kodiyalam, R. K. Kalia, H. Kikuchi, A. Nakano, F. Shimojo, and P. Vashishta, Phys. Rev. Lett. **86**(1), 55 (2001)
68. S. Kodiyalam, R. K. Kalia, A. Nakano, and P. Vashishta, Phys. Rev. Lett. **93**(20), 203401 (2004)
69. N. J. Lee, R. K. Kalia, A. Nakano, and P. Vashishta, Appl. Phys. Lett. **89**, 093101 (2006)
70 S. E. Baltazar, A. H. Romero, J. K. Rodríguez-López, H. Terrones, and R. Martoňák, Comput. Mater. Sci. **37**(24), 526 (2006)
71. D. Frenkel and B. Smit, Understanding Molecular Simulation, Academic Press, New York, (2002)
72. H. Sowa, Solid State Sci. **7**, 1384 (2005)
73. S. H. Tolbert and A. P. Alivisatos, Science **265**(5170), 373 (1994)
74. C. C. Chen, A. B. Herhold, C. S. Johnson, and A. P. Alivisatos, Science **276**, 398 (1997)
75. D. Zaziski, S. Prilliman, E. C. Scher, M. Casula, J. Wickham, S. M. Clark, and A. P. Alivisatos, Nano Lett. **4**(5), 943 (2004)
76. M. Grünwald, E. Rabani, and C. Dellago, Phys. Rev. Lett. **96**, 255701 (2006)
77. J. J. Shiang, A. V. Kadavanich, R. K. Grubbs, and A. P. Alivisatos, J. Phys. Chem. **99**, 17417 (1995)
78. J. R. McBride, T. C. Kippeny, S. J. Pennycook, and S. J. Rosenthal, Nano Lett. **4**(7), 1279 (2004)
79. F. Shimojo, S. Kodiyalam, I. Ebbsjö, R. K. Kalia, A. Nakano, and P. Vashishta, Phys. Rev. B **70**, 184111 (2004)
80. D. Zahn, Y. Grin, and S. Leoni, Phys. Rev. B **72**, 064110 (2005)
81. K. Sarasamak, A. J. Kulkarni, M. Zhou, and S. Limpijumnong, Phys. Rev. B **77**, 024104 (2008)
82. A. J. Kulkarni, M. Zhou, K. Sarasamak, and S. Limpijumnong, Phys. Rev. Lett. **97**, 105502 (2006)

CHAPTER 4

NONEQUILIBRIUM MOLECULAR DYNAMICS AND MULTISCALE MODELING OF HEAT CONDUCTION IN SOLIDS

SIMON P.A. GILL

Department of Engineering, University of Leicester, Leicester, UK, e-mail: spg3@le.ac.uk

Abstract: Modeling methodologies for conducting concurrent multiscale simulations in solids at finite temperature are reviewed. The application of such models to the simulation of inhomogeneous thermal problems is of particular interest. Firstly, the basic methods for temperature control of molecular dynamics (MD) simulations are presented. The derivation of fundamental thermophysical properties from the quantum model of phonons is then outlined, and the relevance of classical MD simulation to heat transport phenomena discussed. Progress in fully atomistic modeling of heat transport is reviewed in relation to nonequilibrium molecular dynamics (NEMD) simulation. Different approaches to isothermal finite temperature multiscale modeling are presented. Equations of motion for coarse-grained dynamics are derived and subject to comment. The further requirements of conservation of thermal energy and the approaches to the transport of heat in non-isothermal multiscale simulations are discussed. Recent progress in this relatively new area of modeling is reported and areas for further work identified

Keywords: Molecular dynamics, Nonequilibrium, Heat conduction, Thermal conductivity, Thermostats, Multiscale, Atomistic-continuum simulation, Coarse-graining, Crystalline solids

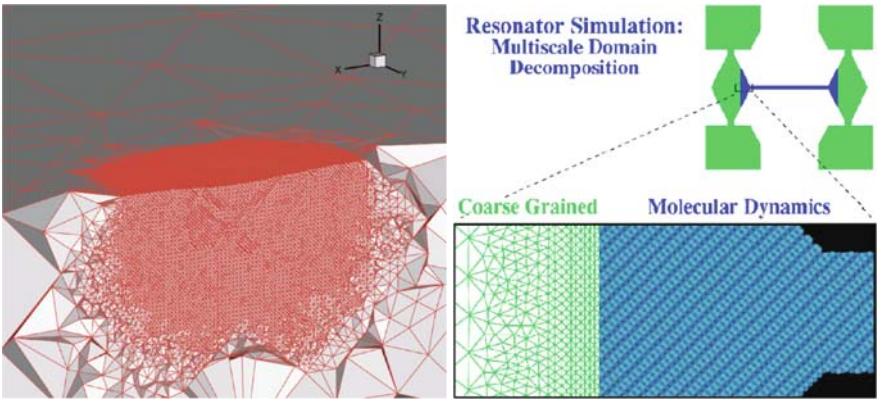
4.1. INTRODUCTION

Typically molecular dynamics (MD) simulations are conducted under equilibrium (or near equilibrium) conditions of constant energy (E) or temperature (T), constant atomic number (N) or chemical potential (μ), and constant volume (V) or pressure (P). The most widely adopted statistical sampling ensembles [1] are the microcanonical ensemble (constant NVE), the canonical ensemble (constant NVT) and the isobaric-isothermal ensemble (NPT) for solids. The grand canonical ensemble (constant μVT) is most commonly applied to fluids, where the number of particles is allowed to fluctuate. There has been a large body of work dedicated to the

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 85–134.
DOI 10.1007/978-1-4020-9785-0_4,

design and implementation of algorithms to generate these ensembles, particularly for thermostatic (constant T) and bariometric (constant P) control. However, there is increasing interest in conducting MD simulations which do not fall within the classification of these classical ensembles. These broadly fall in to the category of nonequilibrium molecular dynamics (NEMD). A simple example is the imposition of a temperature gradient across a sample, for which different temperatures at the boundaries are prescribed [2]. A more complex example is a system that has work done to it during the simulation in the form of a mechanical deformation (such as crack growth [3, 4], nanoindentation [5] or particle impact [6]) or direct thermal excitation (such as laser annealing [7]). In these cases the temperature, pressure (or a more complex stress state), volume and energy of the system vary spatially and temporally during the simulation in a complex way. Ideally the boundary conditions would be applied at a considerable distance from the region of interest to minimize their effect on its dynamics. However, the size of an atomistic system that can be considered is limited by finite computational resources due to their large numbers of degrees of freedom. Therefore most simulations can "see" their boundaries during the course of a simulation and consequently be affected by them. In conventional MD simulations these boundary conditions are a compromise between reality and practicality. In recent years a number of efforts have been undertaken to increase the size of the system under consideration by representing the far field by a coarse-grained (CG) region with a greatly reduced number of degrees of freedom. The prescribed conditions on the remote boundaries of the continuum therefore have less influence on the dynamics of the fully resolved (atomistic) region. A number of successful efforts have been made in this regard for simulation at zero temperature (for a review see [8]). The most notable of these multiscale methods for quasi-static problems being the quasi-continuum method [9]. Figure 4-1a shows a cross-section through a CG substrate subject to deformation under a nanoindenter (not shown). The area beneath the indenter has full atomic resolution where the material response is expected to be highly non-linear (i.e. plasticity due to nucleation of dislocations). This is embedded in a consistent finite element (nonlinear) elastic mesh which models the response of the far-field. Wave propagation at zero temperature has been modeled using coarse-grained molecular dynamics (CGMD) [10]. This is useful if the dynamic response of a body is of interest. Figure 4-1b shows a CGMD model of a nanoelectromechanical system (NEMS) silicon microresonator. Full atomic resolution is retained in the narrow region of the resonator to capture finite size effects due to wave scattering from interfaces.

Since these pioneering early works, there have been a few attempts to extend this type of hybrid atomistic-continuum method to finite temperature. This is overall a more challenging problem for which a number of different approaches have recently been proposed. These methods are divided into two categories. Isothermal multiscale models are considered in Section 4.4. Multiscale models for inhomogeneous thermal problems are the subject of Section 4.5, where heat transport and conservation of thermal energy are a requirement. For this purpose, current atomistic methods for modeling heat transport and determining coefficients of thermal conductivity are reviewed in Section 4.3. The basics of MD simulation are briefly



(a) (b)

Figure 4-1. Examples of two different zero temperature coarse-graining methods. (**a**) Quasi-continuum model for simulation of nanoindentation of Au showing a cross section of a fine-scale atomistic region embedded within a coarse-scale finite element mesh (Reprinted (Fig 6, pg 226102-3) with permission from [Knap J, Ortiz M, Phys Rev Lett, 90, 2003]. Copyright by the American Physical Society. http://link.aps.org/abstract/PRL/v90/p226102). (**b**) Schematic of a coarse-grained molecular dynamics (CGMD) simulation of a NEMS silicon microresonator. The CG region comprises most of the volume, but the MD region contains most of the simulated degrees of freedom (Reprinted (Fig 1, pg 144104-3) with permission from [Rudd RE, Broughton JQ, Phys Rev B, 72, 2005]. Copyright by the American Physical Society. http://link.aps.org/abstract/PRB/v72/p144104)

documented in Section 4.2 to introduce the nomenclature, along with definitions of the relevant thermophysical properties. Such quantities are necessarily derived from a fundamental quantum model of phonon heat transport to demonstrate the relevance of classical MD to this field.

4.2. MOLECULAR DYNAMICS AND ITS APPLICABILITY TO THE SIMULATION OF HEAT TRANSPORT

Firstly some of the fundamentals of MD simulation and temperature control are reviewed. Then the quantum model of phonon-phonon heat transport in a crystalline solid is introduced. This raises issues about the limitations of classical MD and the valid range of application of such atomistic methods is discussed.

4.2.1. Introduction to Equilibrium MD

The basic principles of MD simulation are well-documented [1, 13, 14] and only briefly summarized here. This is a classical method in which particles interact via Newton's second law

$$m_i \frac{d^2 \underline{x}_i}{dt^2} = -\frac{\partial V}{d\underline{x}_i} \tag{4-1}$$

where m_i and $\underline{x}_i$ are the mass and position vector of the ith particle and the total potential energy $V(\underline{x}_i) = \sum_i \sum_{j>i} \phi(r_{ij})$ is the sum of all the interatomic potentials, $\phi(r_{ij})$, where $r_{ij} = \left|\underline{x}_i - \underline{x}_j\right|$ is the interatomic separation. This expression is for simple pairwise atomic interactions which is reasonable for most solid metals and fluids. For covalent solids, such as silicon or carbon, an additional three-body term is necessary to include the bond angle dependence of the potential. More information on material specific interatomic potentials can be found elsewhere [14, 15]. Note that interatomic potentials are always an approximation. They are determined either by fitting the function to reproduce known bulk properties of the material (such as equilibrium lattice spacing, crystal structure, elastic modulus etc) or matching the potential to the results of first-principles calculations. The reliability of such potentials to model complex non-linear phenomena such as defect structures/energies and heat transport coefficients cannot always be taken for granted.

One of the fundamental restrictions on the use of MD is the very small time step. This is determined by the frequency of atomic vibrations in the solid and is typically of the order of femtoseconds. Efficient numerical integration of (4-1) is therefore very important. One of the simplest and most accurate methods is the second-order Verlet algorithm [16], although higher order algorithms such as Gear predictor-corrector [1] and Runge-Kutta [17] are sometimes used. Constant temperature MD simulations are usually initiated at a given temperature by randomly assigning the atoms velocities from the classical Maxwell-Boltzmann distribution subject to zero net momentum. Equipartition theory states that the total energy of a system is equally divided between kinetic and potential energy in the classical regime. Thermal energy is associated with the vibrational (kinetic) energy of atoms such that in a system of N atoms the absolute temperature, T, is defined by

$$^{d}/_{2} N k_B T = {}^{1}/_{2} \sum_{i=1}^{N} m_i \left|\dot{\underline{x}}_i - \left\langle \dot{\underline{x}}_i \right\rangle\right|^2 \tag{4-2}$$

where d is the dimension (1, 2 or 3), k_B is Boltzmann's constant, the dot denotes differentiation with respect to time and the time-averaged velocity of particle i, $\left\langle \dot{\underline{x}}_i \right\rangle$, is subtracted from the total velocity to obtain the vibrational component. It is argued that temperature can only be resolved at a length scale comparable with the mean free path [18] but it is common to define the *kinetic temperature* of an atom from (4-2) as

$$T_i = \frac{m_i \left|\dot{\underline{x}}_i\right|^2}{k_B d} \tag{4-3}$$

where the net velocity of atoms in solids is typically negligible. (Note that this is not the case in fluids where the determination of (4-2) can be problematic. For this reason the *configurational temperature* [19, 20] is often used. This only requires knowledge of the spatial distribution of atoms, not their velocities).

4.2.2. Temperature Control

The system temperature is then regulated by an algorithmic *thermostat*. There are a wide variety of different methods for controlling the temperature of MD simulations. Ideally these should reproduce the canonical Maxwell-Boltzmann distribution but this is not always the case. The first thermostat to produce the canonical ensemble was the stochastic collision method of Andersen [21]. Some other methods, such as velocity exchange [22], also generate the canonical ensemble. Velocity rescaling algorithms [22] are not generally canonical, although this is not strictly always the case [23]. The method of Gaussian constraint [24, 25] produces the iso-kinetic ensemble which is not physically realized. Three of the most popular thermostatting methods are summarised here.

(i) The Nosé-Hoover thermostat. This is a deterministic thermostat which maintains the average temperature of an atomic ensemble at a target value [26, 27]. This is widely used for constant temperature dynamical simulations due to its symplectic, volume conserving, time-reversible Hamiltonian structure which render it beneficial numerical properties. The motion of a thermostatted particle is described by

$$\begin{aligned} m_i \ddot{\underline{x}}_i &= -\frac{\partial V}{\partial \underline{x}_i} - \xi m_i \dot{\underline{x}}_i \\ Q\dot{\xi} &= \frac{\overline{T}(t)}{T_c} - 1 \end{aligned} \tag{4-4}$$

where Q is a (constant) thermal mass, ξ is a thermostatting variable which drives the instantaneous mean temperature, $\overline{T}(t) = {}^1/_{\mathrm{N}} \sum_{i=1}^{N} T_i$ towards the prescribed target temperature T_c. This is a global thermostat in that it enforces an ensemble of particles to maintain an average kinetic energy over time. The choice of the thermal mass is important to ensure sufficient thermal coupling and avoid thermal oscillations [28]. A variant on this thermostat is the Nosé-Hoover chain [29] whereby the thermostatting variable is thermostatted by another thermostatting variable and so on. This is applicable for small and stiff systems as it improves ergodicity.

(ii) The Langevin thermostat. This is a stochastic thermostat which adds a random force to the particle motion along with an appropriate damping term such that the motion of a thermostatted particle is governed by [30]

$$m_i \ddot{\underline{x}}_i = -\frac{\partial V}{d\underline{x}_i} - \gamma m_i \dot{\underline{x}}_i + R\underline{f} \tag{4-5}$$

where γ is a damping coefficient, $-1 \leq R \leq 1$ is a uniformly distributed random variable and $f_n = \sqrt{\frac{6\gamma m_i T_c}{\Delta t}}$ is the magnitude of each component ($n = x$, y or z) of the stochastic force $\underline{f}$ for a time step Δt. The Langevin thermostat is described as a local thermostat as the target temperature can be specified for each atom.

(iii) The Berendsen thermostat. This is another local thermostat, although unlike the Langevin thermostat it is deterministic. In this case the velocities are rescaled at each time step such that the rate of change of temperature is proportional to the deviation in the temperature such that [31]

$$\frac{d\overline{T}(t)}{dt} = \frac{(T_c - \overline{T}(t))}{\tau} \tag{4-6}$$

where τ is a characteristic relaxation time. The resulting scaling factor for the velocities is $\sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_c}{\overline{T}(t)} - 1\right)}$. Like most velocity rescaling algorithms, the Berendsen thermostat does not rigorously generate the canonical ensemble. However, it is still widely used due to its simplicity.

Other specialized thermostats include Recursive Multiple Thermostats [32] for problems with inherently different timescales (particularly biomolecules), configurational temperature thermostats [33] (mainly advantageous for fluids), hybrid deterministic/stochastic thermostats such as Hoover-Langevin [34] (with random noise to improve ergodicity), thermostats based on the generalized Langevin equation [35] (to avoid wave scattering at interfaces) and temperature regulated thermostats [36] (for maintaining equilibrium with minimal intervention). However, it should be noted that any thermostatting algorithm will corrupt the dynamics of the system. Whether this is important depends on the nature of the problem under consideration. For long time sampling of equilibrium thermodynamic quantities such as the thermal expansion coefficient this is not problematic. The consequences for nonequilibrium processes such as heat transport are an issue however. This will be discussed further in the context of nonequilibrium MD simulation in Section 4.3.2. The remainder of this section considers the advantages and disadvantages of using MD to model heat transport in crystal lattices.

4.2.3. Lattice Vibrations

Thermal energy in a solid is due to atomic vibrations. In classical mechanics heat conduction is via the propagation and interaction of normal modes of atomic vibration. In most real crystals there are two mode types: the *acoustic* branch (lower frequencies) and the *optical* branch (higher frequencies). In practice heat transport is predominantly due to the acoustic modes and the optical modes are often ignored. Each branch consists of longitudinal and transverse mode branches. The vibration is in the direction of wave propagation for longitudinal waves and perpendicular to it for transverse waves. For a d-dimensional system ($d = 1$, 2 or 3) there is one longitudinal branch and (d–1) transverse branches. Hence in one-dimension there are no transverse waves but in three-dimensions there are two transverse branches. These differences introduce a strong dependence on the dimensionality to the problem. Using mode coupling theory, Lepri et al. [37] showed that the thermal conductivity in a system of typical size L scaled as $L^{2/5}$ in one-dimension, $\ln(L)$ in two-dimensions and was independent of L in three-dimensions. This size effect in

lower dimensions is due to the absence of interaction between the transverse modes. This has consequences for the simulation of systems in reduced dimensions, either due to the constraints of the model or due to the constraints of the system, e.g carbon nanotubes are low dimensional systems.

The relationship between the wave frequency and the mode number is called the *dispersion relation*. This is often written as $\omega(\underline{K})$, where $\underline{K}$ is the wavevector. For a one-dimensional solid, the wavevector is a scalar known as the wavenumber, related to the wavelength λ by $K = 2\pi/\lambda$. For a chain of N atoms with lattice spacing a_0 one has $K = n\pi/L$ where n is the wave mode (an integer) and $L = Na_0$ is the length of the chain. One can see that the crystal size and lattice spacing place upper and lower limits on the wave mode, $1 \leq n \leq N$, and the related wavelength, $2a_0 \leq \lambda \leq 2L$. This system size dependence is not only important in determining the thermal properties of nanostructures but also in determining the accuracy of MD simulations (naturally of constrained size). Wave scattering at interfaces (real or not) known as the *Kapitza effect* (or phonon scattering) reduces the thermal conductivity due to the local decorrelation of atomic vibrations.

4.2.4. The Quantum Model of Phonon Heat Transport

The true quantum nature of atomic vibrations is not captured by classical models such as MD (see [38, 39] for recent reviews). It is useful to give a brief introduction to the quantum model to appreciate the shortcomings of using classical methods. Quantum theory [40] states that the energy of wave modes can only have certain discrete values. These packets of vibrational energy are known as *phonons*. The energy of the nth mode is given by

$$E_n = \hbar\omega_n \tag{4-7}$$

where $\hbar$ is Planck's constant and $\omega_n = \omega(K_n)$ is the frequency of the nth mode (in radians per second). Phonons are categorized in particle physics as belonging to the family of "particles" known as bosons. These are "particles" that do not obey Pauli's exclusion principle (such that there is no constraint on their number at a particular point in time or space). At thermal equilibrium, the probability of finding N_n mode-n phonons in the canonical ensemble is given by the Boltzmann factor

$$P(N_n) = \frac{e^{-\beta N_n E_n}}{Z_n} \tag{4-8}$$

where $\beta = 1/k_B T$ and the normalizing factor (for an infinite crystal)

$$Z_n = \sum_{N_n=0}^{\infty} e^{-\beta N_n E_n} = \frac{1}{1 - e^{-\beta E_n}} \tag{4-9}$$

ensures unit total probability and is known as the *partition function* for the nth mode. The average number of mode-n phonons is therefore given by

$$\langle N_n \rangle = \sum_{N_n=0}^{\infty} N_n P(N_n) = \frac{1}{e^{\beta E_n} - 1} = f(E_n) \tag{4-10}$$

which is known as the *Bose-Einstein* distribution [41]. The total internal energy of a (one-dimensional) solid is therefore

$$U = \sum_n E_n f(E_n). \tag{4-11}$$

For isotropic systems of infinite size the number of vibrational modes becomes infinite. The above sum can therefore be expressed as an integral, which is most conveniently written in terms of the (continuous) frequency as

$$\sum_n \rightarrow \int g(\omega) d\omega \tag{4-12}$$

where $g(\omega)$ is known as the *density of states*. It is the number of states with a frequency in the range ω and $\omega + d\omega$, and is derived from the dispersion relation [40]. Thus (4-11) is often written in continuous form as

$$U = \int E(\omega) f(\omega,T) g(\omega) d\omega \tag{4-13}$$

where $E(\omega) = \hbar\omega$ etc. The (volumetric) specific heat per atom is the thermal energy required to raise its temperature by 1 K such that

$$C = \frac{1}{N}\frac{\partial U}{\partial T} = \frac{1}{N}\int E(\omega)\frac{\partial f}{\partial T} g(\omega) d\omega. \tag{4-14}$$

The thermal conductivity is given by

$$k = \frac{1}{Nd}\frac{\partial}{\partial T}\int E(\omega) f(\omega,T) g(\omega) v(\omega) \Lambda(\omega) d\omega \tag{4-15}$$

where the phonon transmission velocity (or *group velocity*)

$$v(\omega) = \frac{d\omega}{dK} \tag{4-16}$$

is given by the dispersion relation, and $\Lambda(\omega)$ is the *phonon mean free path*. This is the average distance a phonon travels before it interacts with another phonon (or an

interface). Heat transport is via the (anharmonic) interaction of phonons, sometimes referred to as *Umklapp scattering* or the U-process [40]. Note that if the velocity and mean free path are assumed to be constant then, using (4-14), (4-15) can be simplified to the familiar formula for (three-dimensional) thermal conductivity

$$k = {}^1/_3 C v \Lambda. \tag{4-17}$$

Transport coefficients, such as the thermal conductivity, arise from the summation of a number of complex interaction processes throughout the entire lattice. They are not readily approximated analytically, but can be determined by simulation or experiment. This is addressed in Section 4.3. However, simple approximations about the nature of wave propagation in a solid can yield useful analytical predictions for other thermomechanical properties, such as the heat capacity. Two simple models are briefly reviewed here.

(i) The Debye model. Debye [41] made the simplifying assumption that the dispersion relation is given by $\omega = cK$ where c is constant. Although this is not completely true (even for an ideal harmonic crystal) it is often a reasonable approximation for the lower frequency modes. We can see from (4-16) that $v(\omega) = c$ so c is the speed of sound in the crystal. In three-dimensions the summation becomes

$$U = 3 \sum_{n_x} \sum_{n_y} \sum_{n_z} E_n f(E_n) \tag{4-18}$$

where $n = \sqrt{n_x^2 + n_y^2 + n_z^2}$ is the total magnitude of its three components and the factor of 3 appears because of the existence of three phonon branches. The group velocity of the longitudinal and two transverse branches is assumed to be the same. Debye approximated this summation by an integral over an eighth of a sphere such that

$$\sum_{n_x} \sum_{n_y} \sum_{n_z} \rightarrow \frac{\pi}{2} \int n^2 dn = \int g(\omega) d\omega \tag{4-19}$$

The dispersion relation is $\omega = cn\pi/L$ so the density of states is $g(\omega) = \frac{\pi n^2}{2} \frac{dn}{d\omega} = \frac{V_0 \omega 2}{2\pi^2 c^3}$, where $V_0 = L^3$ is the volume of the crystal. From (4-10) and (4-13) the internal energy of the crystal per atom is

$$\frac{U_{\text{Debye}}}{N} = 9(k_B T) \left(\frac{T}{\theta_D} \right)^3 \int_0^{\theta_D/T} \frac{x^3}{e^x - 1} dx \tag{4-20}$$

where $\theta_D = \sqrt{\frac{6\pi^2}{\Omega}} \frac{\hbar c}{k_B}$ is the *Debye temperature*, $\Omega = V_0/N$ is the atomic volume and $x = \hbar\omega/k_B T$. Alternatively the Debye temperature is expressed in terms of the *Debye*

frequency, ω_D, where $k_B\theta_D = \hbar\omega_D$. The Debye approximation for the specific heat capacity can therefore be calculated from (4-14) and (4-20) such that

$$C_{\text{Debye}} = 9k_B\left(\frac{T}{\theta_D}\right)^3 \int_0^{\theta_D/T} \frac{x^4 e^x}{(e^x-1)^2}dx \tag{4-21}$$

(ii) The Einstein model. This considers the atoms to be non-interacting simple quantum harmonic oscillators vibrating at a fundamental frequency, ω_0. Due to this decoupling, the energy levels of the oscillator

$$E_n = \hbar\omega_0(n - {}^1/_2) \tag{4-22}$$

are similar to be energy of the Debye modes except for the introduction of an additional non-zero reference energy ${}^1/_2\hbar\omega_0$, known as the zero-point energy. The Einstein density of states is therefore $g(\omega) = \delta(\omega - \omega_0)$, where $\delta(\omega)$ is the Delta function. Hence (4-13) can be written as

$$\frac{U_{\text{Einstein}}}{N} = {}^1/_2\hbar\omega_0 + \frac{\hbar\omega_0}{e^{\beta\hbar\omega_0}-1} \tag{4-23}$$

for a one-dimensional oscillator. In three-dimensions this can be expressed as

$$\frac{U_{\text{Einstein}}}{N} = \sum_{k=1}^{3} \frac{\hbar\omega_k}{2}\coth\left(\frac{\beta\hbar\omega_k}{2}\right) \tag{4-24}$$

where ω_k are the three fundamental frequencies of the three-dimensional lattice. The atomic heat capacity also has a simple analytical expression from (4-14)

$$C_{\text{Einstein}} = k_B\sum_{k=1}^{3}\left(\frac{\beta\hbar\omega_k}{2}\right)^2 \frac{1}{\sinh^2\left(\beta\hbar\omega_k/2\right)} \tag{4-25}$$

Using the *local harmonic approximation*, the fundamental frequencies of a lattice are given by the determinant of the *dynamical matrix*

$$\left| m_i\lambda^2\underline{I} - \frac{\partial^2 V}{\partial\underline{x}_i\partial\underline{x}_i}\right| = 0 \tag{4-26}$$

where the ω_k are the three eigenvalues λ of this 3×3 matrix for atom i and $\underline{I}$ is the identity matrix. From (4-26) is it clear that the local eigenfrequencies can depend

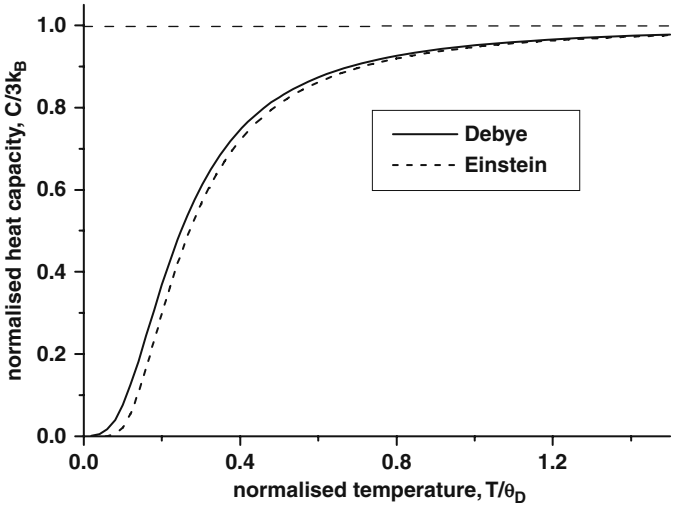


Figure 4-2. Debye (4-21) vs Einstein (4-25) prediction for the atomic heat capacity as a function of temperature. The Debye model is accurate at low temperatures whereas the Einstein model is not. The models are indistinguishable at high temperatures. The classical limit is shown as a *dashed line*. It is generally accepted that classical MD is valid above the Debye temperature ($T/\theta_D > 1$)

on the potential energy of the body, $V(\{\underline{x}_i\})$, and hence change locally in the vicinity of defects and globally if the body is elastically deformed. For an undeformed, isotropic solid we can associate the fundamental frequencies with the highest sustainable frequency in the lattice. This allows the predictions of the Debye and Einstein models to be compared. Figure 4-2 shows the heat capacity as a function of temperature. It is clear that both models agree at high temperature. However there is a significant departure at low temperatures. Experimentally it has been shown that the Debye model is valid in this regime but the Einstein model is not. However, as we shall see in Section 4.4.1, the Einstein model is of benefit to multiscale simulation as it is reasonably good at higher temperatures, analytically tractable and relates the thermophysical state to the local dynamic properties of the matrix.

4.2.5. The Classical Limit

The previous subsection considered the quantum nature of lattice vibrations. This is not accounted for in classical Newtonian MD. This subsection considers under what conditions the two models are compatible. Compare the predictions of the quantum models in Figure 4-2 with the expectations for a classical (non-quantized) system. In a classical system energy is equally partitioned between the kinetic and potential energy of the atoms (on average). In crystalline systems each atom has 3 translational degrees of freedom. Classical models of temperature state that the kinetic energy of each degree of freedom is $\frac{1}{2}k_BT$. Hence the total kinetic energy per an atom in three-dimensions is $\frac{3}{2}k_BT$ and similarly for the potential energy. For

N atoms, the total internal energy is therefore $U_{\mathrm{MD}} = 3Nk_BT$ and hence the classical heat capacity per atom is $C_{\mathrm{MD}} = \frac{1}{N}\frac{dU_{\mathrm{MD}}}{dT} = 3k_B$. This is known as the *Dulong-Petit law*. It is easy to see in Figure 4-2 that the quantum heat capacities, given by (4-21) and (4-25), converge to this limit for high temperatures. At the Debye temperature $T = \theta_D$ the quantum heat capacity is $C_{\mathrm{debye}} = 0.95\ C_{\mathrm{MD}}$. For this reason, it is commonly accepted that classical Newtonian dynamics are acceptable for the modeling of heat conduction in insulators above the Debye temperature. In this regime the Dulong-Petit law for heat capacity is very reasonable for most materials. It is still valid for many materials at low temperatures (e.g. silicon) but fails for metals, where the contribution to the heat capacity from electrons becomes significant. Values for the Debye temperature for some common elements are shown in Table 4-1. In total, 30 of the elements have Debye temperatures above room temperature of which only three are not metals: carbon, silicon and germanium. The high value for carbon means that the heat capacity is 25% of the classical value at room temperature, i.e. $C_{\mathrm{debye}} = 0.25\ C_{\mathrm{MD}}$. At high temperatures one can also see that that the Bose-Einstein distribution (4-10) is equivalent to the classical *Maxwell-Boltzmann* distribution such that

$$f(E_n) \approx e^{-E_n/k_BT} \quad \text{for} \quad T > \theta_D \tag{4-27}$$

Table 4-1. Table of Debye temperatures for some common elements [40]. Their position in the list of elements (highest Debye temperature first) is shown

Position	Element	Debye temperature, θ_D
1st	Carbon	2230 K
2nd	Beryllium	1440 K
3rd	Silicon	645 K
4th	Chromium	630 K
8th	Iron	470 K
9th	Nickel	450 K
13th	Aluminium	428 K
20th	Germanium	374 K
35th	Silver	225 K

The validity of classical MD is summarized nicely in Figure 4-3. It is valid over the range of most practical length scales. MD obeys Fourier's law quite closely even under high temperature gradients [42] but the conductivity varies with the extent of the domain and changes locally in the vicinity of defects. The dispersion relation (and hence the density of states and phonon velocities) is strongly modified by the size of the crystal. In general all such features lower the thermal conductivity. This property of MD is an advantage for the simulation of heat transport in nanostructures in that it includes size effects due to the inherent restrictions on the mean free path of energy carriers and scattering at interfaces. One disadvantage is that

it is difficult to remove these size effects if the modeler is interested in the bulk properties. This is discussed further in Section 4.3.3. Another disadvantage is that MD results are only valid in the classical regime (above the Debye temperature). As shown in Figure 4-3, modeling methodologies such as perturbation theory (PT) and the Boltzmann Transport Equation (BTE) can be adopted at these small scales if required. See Sinha and Goodson [43] and Murthy et al. [39] for recent reviews on sub-continuum heat transport. It has recently been proposed that low-temperature quantum effects due to discrete phonons can be incorporated into classical MD simulations using appropriate quantum thermostats [44, 45]. Finally, MD is restricted to the analysis of heat transport due to phonon-phonon interactions only. This is the dominant transport process in insulators but not in metals. Metals are briefly discussed in the next subsection. MD is typically employed to model heat transport

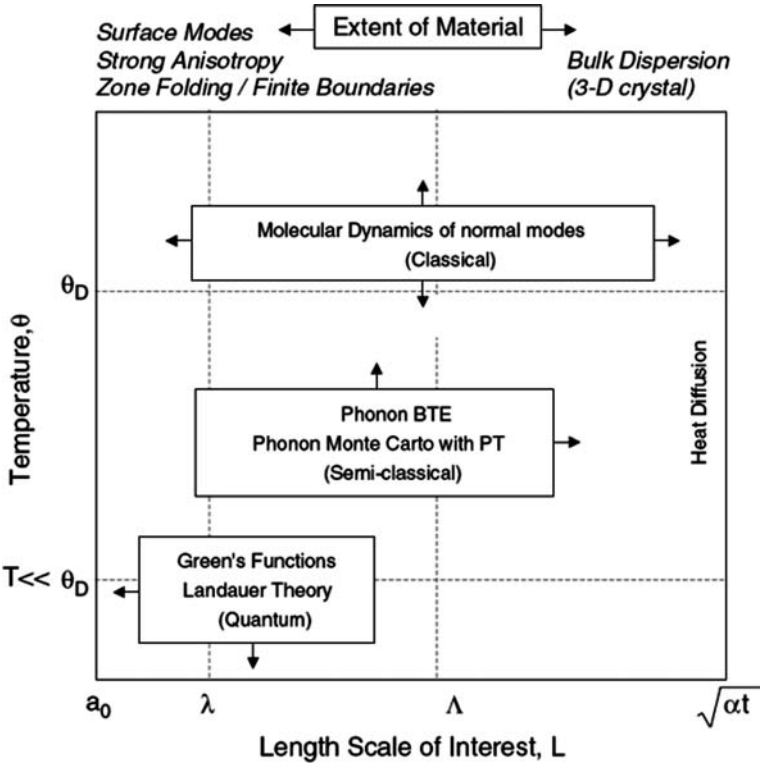


Figure 4-3. The appropriate thermal modeling methodology for insulators depends on the dimension of the system and the temperature. At dimensions comparable to the phonon wavelength λ and temperatures much smaller than the Debye temperature θ_D (*bottom left*), quantum mechanical nature is strongly manifest. At larger dimensions approaching the phonon mean free path Λ and room temperatures, a semi-classical approach is more pragmatic such as perturbation theory (PT) or the Boltzmann Transport Equation (BTE). At temperatures above θ_D the classical physics of MD is applicable. The diffusion length $\sqrt{\alpha t}$ is given for the sake of completeness where α is the thermal diffusivity t is the phonon relaxation time t (Reprinted from [Sinha S, Goodson KE (2006) Thermal conduction in sub-100nm transistors. Microelectronics Journal 37:1148] with permission from Elsevier)

in carbon (important for nanotubes, fullerenes, graphite etc) and silicon and germanium (which are important materials in transistors and many nanostructures [46]). Results for these materials are reviewed in Section 4.3.3.

4.2.6. Heat Transport in Metals

Table 4-2 shows a comparison between values for the thermal conductivity of different elemental solids from experiment and as determined from MD simulation (see Section 4.3.2 for details). The simulated values for the metallic elements are one or two order of magnitude in difference from the experimental values, even though the temperatures are within or close to the classical regime. The values for silicon are within the margin of error, and the values for carbon are of the same magnitude but the difference is not inconsiderable as the temperature is well below the Debye value. The dramatic difference between these materials is because the majority of heat transport in metals is due to electrons which are not accounted for in MD simulations, whereas electronic heat transfer is negligible for insulators. Electrons also contribute to the heat capacity at low temperatures [40].

Scattering effects due to the dominant electron-phonon interactions in metals have been included for modeling electrothermal transport in transistors [43, 46]. There are only a few MD studies of heat transport in metals (see [48, 49] and references therein). These assume that electron-phonon interactions are the dominant heat transport process and that phonon-phonon (ballistic) interactions can be ignored if the electrical conductivity is high [48]. The conductance due to phonon-phonon interactions is determined from an MD simulation. The dominant phonon-electron contribution is then determined from estimation of the phonon-electron mean free path [48] or Wiedemann-Franz law [49] where the thermal conductivity associated with conductance electrons is deduced from the electrical conductivity. The Wiedemann-Franz law states that, for metals at not too low temperatures, the ratio

Table 4-2. Thermal conductivities for various elements (W/mK). Metals use embedded atom method (EAM) potential and silicon and carbon use Stilinger-Weber (SW) potential. The estimated errors on the simulated conductivities are due to extrapolating the finite size results of the simulation to an infinite crystal. The allotrope employed for carbon is diamond

Element	Experimental	Simulated	Temperature (K)
Silver[a]	420	1.2	300
Aluminium[a]	242	3.6	300
Gold[a]	309	0.4	300
Copper[a]	405	4.3	300
Nickel[a]	86	6.5	300
Silicon[b]	120	119±40	500
	50	65±16	1000
Carbon[b]	400	573±60	1000

[a]From reference [47].
[b]From reference [2].

of thermal (k) to electrical (σ) conductivity is proportional to the temperature such that

$$\frac{k}{\sigma} = \frac{\pi^2}{3}\left(\frac{k_B}{e}\right)^2 T = LT \tag{4-28}$$

where e is the electron charge and the constant $L = 2.45 \times 10^{-8}$ W Ohm K^{-2} is known as the *Lorenz number*.

4.3. NONEQUILIBRIUM MOLECULAR DYNAMICS

The previous section considered the practical requirements and limitations of conducting constant temperature (equilibrium) MD simulations. The extension of these principles to the investigation of non-isothermal *nonequilibrium molecular dynamics* (NEMD) is the subject of this section. Away from equilibrium, the flow of heat within a body needs to be conserved such that

$$\rho c_p \frac{\partial T}{\partial t} = -\nabla j \tag{4-29}$$

where ρ is the material density (kg m^{-3}), c_p is the specific heat capacity (J kg^{-1} K^{-1}) and j is the heat flux (W m^{-2}). Fourier proposed a simple constitutive relationship between the heat flux and temperature gradient such that

$$j_q = -\sum_{r=1}^{3} k_{qr} \frac{\partial T}{\partial x_r} \tag{4-30}$$

where k_{qr} is the three-dimensional thermal conductivity tensor (W m^{-1} K^{-1}). For an isotropic solid $k_{qr} = k\delta_{qr}$ where k is the isotropic thermal conductivity and δ_{qr} is the identity tensor. Equations (4-39) and (4-30) then give the classical heat conduction equation

$$\rho c_p \frac{\partial T}{\partial t} = \nabla(k \nabla T) \tag{4-31}$$

where $k(T)$ is often a function of temperature.

There are two widely used methods for the calculation of the (ballistic or phonon) thermal conductivity (of insulators) from atomistic simulation in the classical regime. These methods are discussed in some detail by Schelling et al. [2] and hence are only briefly repeated here.

4.3.1. The Green-Kubo Method

Unlike fluids, the transport of heat in solids is derived from correlations in the motion of atoms in a crystal lattice. The Green-Kubo relations use the fluctuation-dissipation theorem to derive an exact mathematical expression for the thermal conductivity which can be determined from isothermal equilibrium MD simulations. The thermal conductivity is given by a time integral over the equilibrium flux autocorrelation function such that

$$k_{qr} = \frac{1}{V_0 k_B T^2} \int_0^\infty \langle j_q(t) j_r 0) \rangle dt \tag{4-32}$$

where V_0 is the volume of the system and the angular brackets indicate an ensemble average (over the volume of the sample) [38]. Schelling et al. [2] give a discretized version of this equation for atomistic calculation, and a method for calculating the instantaneous heat flux between atoms from the interatomic potential. Webb et al. [50] have gone into the derivation of the heat flux in some detail. Jolley and Gill [42] used a simpler method based on the change in temperature of an atom before and after the time step. This relies on the knowledge of the flux at a given point in the system.

Nonlinear transport coefficients can be calculated from steady state NEMD using a nonlinear version of the Green-Kubo relation called the transient-time correlation function [51]. This is particularly advantageous for fluids where the exceptionally large temperature gradients required for standard NEMD induce large convection currents which make determination of the thermal conductivity problematic.

4.3.2. The Direct Method

In the direct method, a steady state temperature gradient is imposed on the atomistic sample by thermostatting different regions at different temperatures. This is one of the principal examples of NEMD simulations in the literature. The samples usually have one long dimension and a smaller square cross-section. The temperature gradient is imposed along the long dimension. Periodic boundary conditions are used on the sides. Two typical examples of the types of boundary conditions that can be used at the ends are shown in Figure 4-4. In theory the periodic end conditions allow for the existence of longer wavelength modes than the fixed end conditions. However, in practice the thermostats corrupt the lattice dynamics, so the maximum mean free path is typically the distance between the thermostatted regions. A variety of thermostats are often used, such as those discussed in Section 4.2.2. Another variant is to control the rate that heat enters the hot bath and leaves the cold bath regions [50, 52]. This has the same effect but the magnitude of the resulting temperature gradient cannot be determined beforehand and some calibration of the heating rate is required. A velocity rescaling thermostatting algorithm [52] was used in the example shown in Figure 4-5 to avoid the thermostat adding any instantaneous momentum to the system. All thermostats conserve momentum on average, however, and it has been

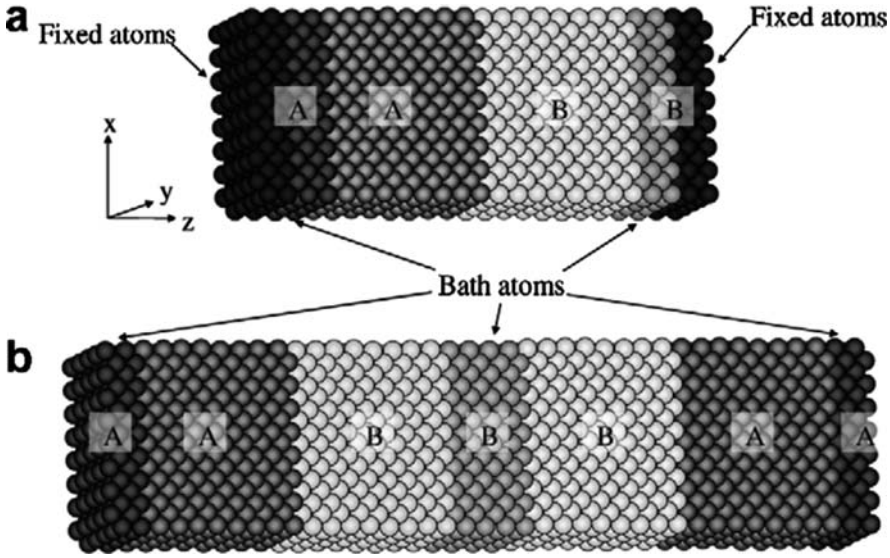


Figure 4-4. Two different boundary conditions for imposing a steady state temperature gradient on a NEMD simulation: (**a**) fixed boundary conditions (four atomic planes are fixed (*black*) and six are used for the hot and cold bath atoms (*dark grey* and *light grey*) but only four are shown here). (**b**) periodic boundary conditions (Reprinted from [Stevens RJ, Zhigilei LV, Norris PM (2007) Effects of temperature and disorder on thermal boundary conductance at solid-solid interfaces: nonequilibrium MD simulations. Int J Heat Mass Transfer 50:3977] with permission from Elsevier)

demonstrated that rigorous momentum control at each time step is not necessary in solids [22].

Upon reaching a steady state, the (time-averaged) variation in temperature across the length of the sample can be determined. A typical example of such a temperature profile is shown for a silicon sample in Figure 4-5 where periodic boundary conditions have been used (such as those shown in Figure 4-4b). The temperature profile is highly nonlinear and not as expected. In the steady state the heat flux must be constant (on average) at all points in the system. For a constant thermal conductivity, Eq. (4-30) predicts that the expected temperature profile is linear between the thermostats. Examination of Figure 4-5 shows that there are two linear regions between the thermostats but there is a sudden change in temperature near the heat sink and source regions themselves. Given the temperature gradient in the linear region(s) the thermal conductivity can be determined from (4-30) as

$$k = j\left(\frac{\partial T}{\partial x}\right)^{-1}. \tag{4-33}$$

This requires a knowledge of the heat flux, j. This is always uniform across the length of the sample in the steady state and can be calculated by several methods (see Section 4.3.1).

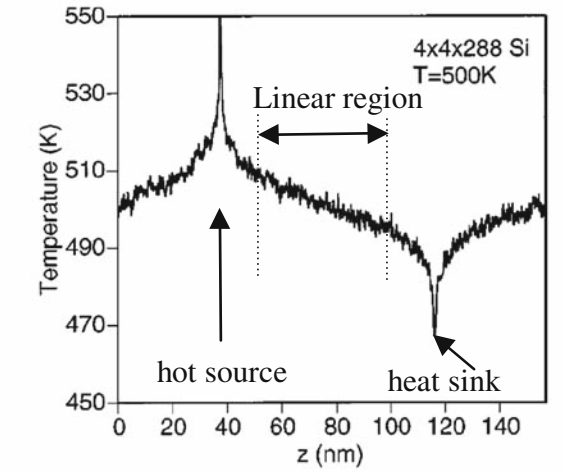


Figure 4-5. Typical temperature profile for an atomic $4 \times 4 \times 288$ silicon system. The heat source is located at $z = 39$ nm, and the heat sink is located at $z = 117$ nm. Within 6 nm of the source and sink, a strong nonlinear temperature profile is always observed. To compute k from (4-30) the temperature gradient is determined from the linear region, which are at least 6 nm away from the heat source and sink (Reprinted (Fig 3, pg 144306-4 and Fig 6, pg 144306-5) with permission from [Schelling PK, Phillpot SR, Keblinski P, Phys Rev B, 65,2002]. Copyright by the American Physical Society. http://link.aps.org/abstract/PRB/v65/p144306)

It is clear from Figure 4-5 that the temperature gradient observed in the linear regions (where Fourier's law (4-30) applies) is very different from the temperature gradient that is applied. The nonlinearity in the temperature profile in the proximity of the thermostatted regions is commonly attributed to phonon scattering at the interface between the thermostatted and non-thermostatted regions. This scattering arises because the thermostatting algorithms necessarily corrupt the dynamics of the thermostatted atoms. As seen in Section 4.3.1, the thermal conductivity is proportional to the long-time flux autocorrelation function. The correlated motions of atoms are sensitive to small disturbances, such as those used by thermostats to regulate temperature. It is clear from (4-31) that if the motions of atoms in a crystal lattice are completely uncorrelated then the thermal conductivity is zero. Thermostats will always reduce the correlation in the motion of atoms and as such this will always reduce the thermal conductivity in these regions. The heat flux is constant in the steady state so (4-31) predicts that a decrease in the thermal conductivity will be associated with an increase in the temperature gradient. This is what is observed in Figure 4-5 in the proximity of the thermostatted regions. This change in the thermal conductivity at an interface is known as the *Kapitza effect* and is observed in all NEMD simulations. It is widely known that the material properties near interfaces are different from those in the bulk, and hence the Kapitza effect is important in nanostructures, where the large proportion of interfaces can dominate their response. However, in NEMD simulations such as that in Figure 4-5, the interface is not real and the observed artificial Kapitza effect

is unphysical and consequently undesirable. This lack of precision in the control of the thermal boundary conditions for NEMD simulations is very important when considering concurrent compatibility conditions with multiscale methodologies (see Section 4.4). Unfortunately it is impossible to avoid completely as temperature control can only be achieved through adjusting the motion of the thermostatted atoms.

Jolley and Gill [42] found that the (artificial) Kapitza effect present in NEMD simulations was significant for deterministic thermostats (such as Nosé-Hoover) but overall the effect was much greater for stochastic thermostats (such as Langevin). This is not surprising as stochastic thermostats rely on uncorrelated random noise to supply the thermal bath which will destroy the correlated motion of the thermostatted atoms. Jolley and Gill [42] also found that the Kaptiza effect was more dominant in lower dimensions and that in three-dimensions the Kapitza effect increased as the cross-section of the sample was reduced below roughly 8×8 (for the Lennard-Jones potential). This is expected to be due to the restricted interaction between the transverse phonon modes in these cases. Thermostatting methods based on memory-kernels (see Section 4.4.3) which utilize information retained offer time and space to minimize disruptions to the correlations in the motion of atoms over potential benefits in minimizing the (artificial) Kapitza effect [35, 54]. However, constant temperature results still show some corruption of the autocorrelation function adjacent to the thermalised region [35]. To the authors knowledge, these techniques have not been applied to NEMD simulations so their advantages in this respect cannot currently be quantified.

A simple pragmatic approach to the artificial Kapitza effect has been proposed by Jolley and Gill [42]. This does not aim to avoid the Kapitza effect, but uses standard thermostats to impose the precise temperature gradient that is desired upon a thermostatted region of the system through a simple control loop. Typical time-averaged temperature profiles for deterministic (Nosé-Hoover) and stochastic (Langevin) thermostatic control are shown for an $8 \times 8 \times 100$ Lennard-Jones solid in Figure 4-6. The aim is to maintain different prescribed temperatures at the boundaries of an (unthermostatted) true dynamics region (TDR) in the centre of the sample. These are defined as $T_0 = 50$ K and $T_M = 40$ K on the left and righthand ends respectively, where $M = 50$ is the number of atomic planes in the TDR. The end conditions are fixed (as in Figure 4-4a). The target temperatures of the thermostatted regions (TR) at the left and righthand ends, T_L and T_R are not known. They are determined during the simulation via a very simple feedback control algorithm

$$Q_T \dot{T}_L = (T_0 - \langle T_0 \rangle) \quad Q_T \dot{T}_R = (T_M - \langle T_M \rangle) \tag{4-34}$$

where $\langle T_0 \rangle$ and $\langle T_M \rangle$ are the atomistic temperatures at the TDR boundaries and the constant Q_T determines the responsiveness of the thermostat. There is a buffer region (BR) between the TRs and the TDR to allow for the nonlinear Kapitza effect. The difference between the temperature profiles for the two thermostats is of interest. Firstly, the temperature within the TR is not constant for both cases. This is not surprising for the Nosé-Hoover thermostat as it is a global thermostat which drives

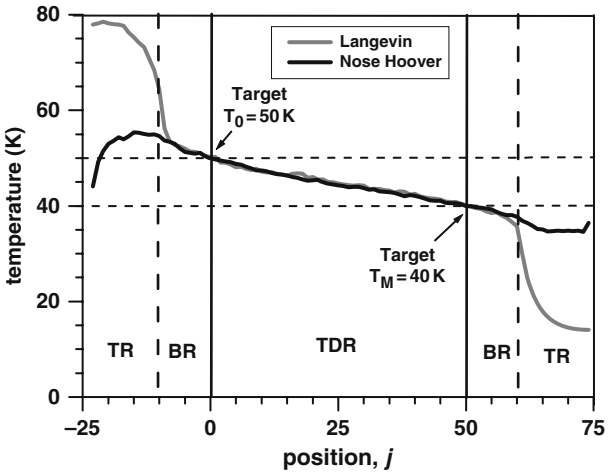


Figure 4-6. Temperature profile across an $8 \times 8 \times 100$ atom Lennard-Jones MD simulation. The temperatures in the thermostatted regions (TR) at the ends are dynamically determined to achieve the target temperatures (50 and 40 K) on the boundaries of the true dynamics region (TDR). The buffer region (BR) allows for non-linear Kapitza effects

the system towards an average temperature. The average temperature is achieved but there is no requirement for the temperature distribution to be uniform. The Langevin thermostat was designed as a local equilibrium thermostat. The target temperature is imposed using (4-5) but (unlike Nosé-Hoover) there is no feedback from the simulation in this algorithm to ensure that the target temperature is achieved. Therefore the Langevin thermostat on its own cannot be guaranteed to maintain a prescribed target temperature in NEMD simulations. However, equation (4-34) introduces a feedback loop which avoids this problem in the above situation. Secondly, note that the temperature difference between the TRs for the Langevin algorithm is much higher than for Nosé-Hoover. As discussed before, this is due to the fact that the artificial Kapitza effect (temperature drop) in the BR is much larger for stochastic thermostats than deterministic ones due to their stronger disruption of the correlations in the atomic vibrations. Both algorithms locally reproduce the canonical ensemble in the TDR and enforce the correct temperatures at the boundaries of the TDR.

One of the general disadvantages of NEMD simulations is that the temperature gradients are very large (of the order of 10^9 K/m). Such a large gradient might be expected to be beyond the applicability of Fourier's law. Although the author has not found this to be the case within the classical MD regime [42] other nonlinear continuum models have been proposed [38, 39] for non-classical heat flow. Temperature gradients of this magnitude are expected across interfaces [53], when two surfaces at different temperatures first come into contact [55] and around hot spots in nanostructured devices [46]. An advantage of the direct method over the Green-Kubo method is that the dependence of the thermal conductivity on temperature can be deduced from the NEMD temperature profile. A linear temperature variation in the thermal conductivity will result in a quadratic variation in the temperature profile [42]. A

quadratic fit to the profile will yield the (extrapolated) thermal conductivity at 0 K as well as the linear temperature dependence term. However, NEMD simulation results will also include a dependence on strain [56]. Although possibly free to expand, the shape of NEMD simulations are constrained to remain cuboidal. Thermal expansion within the body will be non-uniform resulting in a dilational strain gradient along the sample. For highly constrained systems, or large temperature differences, the strain state will have an affect on the calculated thermal conductivity.

A variant on the direct method has been proposed by Terao and Müller-Plathe [57]. In their scheme, the sample is uniformly heated at all points within the body by the addition of a random noise. This random noise is not generated by a thermostat, but arises by deliberately truncating the cut-off distance for the interatomic potential interactions so that the noise is derived from numerical errors. Heat must be removed from the system for it to attain a steady state. A thermostat is used to impose a heat sink at two locations along the periodic sample. The resulting temperature profile is a quadratic variation between the cooling slabs, as shown in Figure 4-7 Although easy to implement, the disadvantage of this approach for solids is that the entire region is thermostatted and hence the dynamics and the resulting transport coefficient will be affected.

Another approach is to monitor the decay of thermal transients in a system. Daly et al. [58] investigated the conductance of superlattices by initially imposing a sinusoidal temperature variation across the sample. Fourier's law (4-31) predicts that such a variation will remain sinusoidal and that its amplitude will decay exponentially with time at a characteristic rate. Hulse et al. [59] applied a transient technique

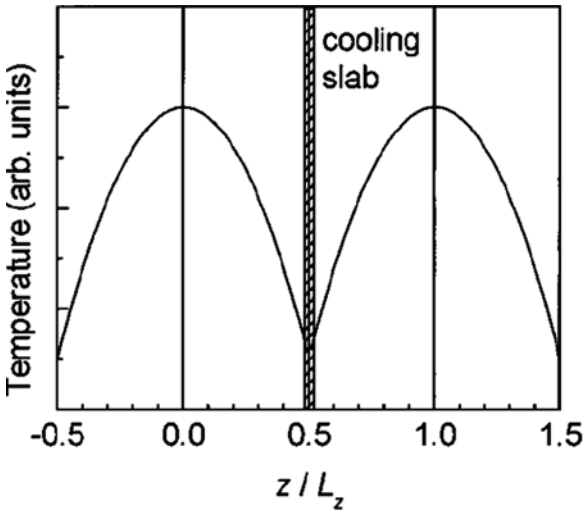


Figure 4-7. Schematic quadratic temperature profile in water calculated under periodic boundary conditions. (Reprinted with permission from [Terao T, Müller-Plathe F (2005) A nonequilibrium MD method for thermal conductivities based on thermal noise. J Chem Phys 122:081103]. Copyright [2005], American Institute of Physics)

to fluids, in which a small region of the simulation was instantaneously heated by velocity rescaling. The thermal conductivity was determined from the rate at which the average temperature in the heated region returned to equilibrium.

In summary, any thermostatting algorithm will cause a reduction in the thermal conductivity in regions within a distance of 3-5 atomic spacings of where it is active. Therefore the thermal conductivity can only be properly determined in regions of true dynamics which are not thermostatted. This suggests that NEMD simulations should only be thermostatted at the boundaries and not in the bulk. Deterministic thermostats tend to be less disruptive to correlated dynamics than stochastic ones. Local thermostats are preferable for prescribing spatially varying thermal boundary conditions, as global thermostats only control the temperature average away from equilibrium, not its distribution. Most thermostats have been designed for simulations at thermal equilibrium. They require some feedback from the actual temperature of the simulation for use in NEMD simulations.

4.3.3. Size Effects

All of the algorithms discussed in the previous subsection are subject to size effects due to phonon scattering at the system (or thermostat) boundaries, and restrictions on the maximum phonon mean free path. In nanostructures, many of these size effects are real and it is desirable that they are observed in representative NEMD simulations. These are discussed at the end of this section. However, it is often desirable to determine the bulk thermal conductivity of a material from a MD simulation. In this case, the finite size thermal conductivity has to be extrapolated up to the bulk value using analytical predictions. The simplest of these is the widely used equation [2]

$$k(L) = k_\infty \left(\frac{L}{L + 4l_\infty} \right) \tag{4-35}$$

where L is the length of the simulation cell, l_∞ is the phonon mean free path in an infinite system and k_∞ is the (bulk) thermal conductivity in an infinite system. This is fitted to results from NEMD simulations of different system sizes in Figure 4-8. The bulk thermal conductivity can be extrapolated from (4-35) by letting $L \to \infty$. Equation (4-35) implies that the thermal conductivity of a sample is prone to size effects when the dimensions fall below 5–10 times the phonon mean free path. For phonons the mean free path is typically 10–100 nm [38].

Such size effects are prevalent in nanostructured materials [38, 60, 61], where the reduction in thermal conductivity can either be beneficial or problematic. For instance, nanolayered structures can be used to produce films with good thermal insulating properties [62]. However, in nanoscale transistors, where Joule heating due to high electric currents is very significant, the increased thermal resistance leads to the development of large adverse temperature gradients and hot spots within the device [46]. NEMD simulations have been employed to investigate a variety of nanostructures, including size effects in argon [18] and silicon [63, 64], nanoconstrictions in silicon [55], nanoporous silicon [65] and silicon nanofilms [63, 66]

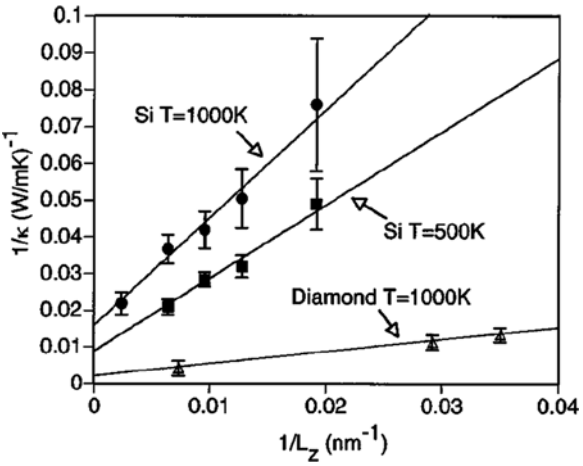


Figure 4-8. System size dependence of thermal conductivity, k, on the sample length, L. Data points shown are calculated from NEMD simulations for Si at $T = 500$ K and $T = 1000$ K and for diamond at $T = 1000$ K for different sample sizes. The bulk conductivities are calculated using (4-35) from lines of best fit. The results are given in Table 4-1 (Reprinted (Fig 3, pg 144306-4 and Fig 6, pg 144306-5) with permission from [Schelling PK, Phillpot SR, Keblinski P, Phys Rev B, 65,2002]. Copyright by the American Physical Society. http://link.aps.org/abstract/PRB/v65/p144306)

and nanowires [67], liquids confined in nanopores [51], quantum wells in semiconductors [68], nanolayered superlattices [58, 62], nanoparticles [49], molecular wires [69] and carbon nanotubes [70–76]. The effect of crystalline defects on phonon scattering [53], such as grain boundaries [77, 78] has also been studied. A significant reduction in the thermal conductivity has been observed in quantum dot superlattices [79] when the quantum dot arrangement in the superlattice have changed from correlated (regular) to completely uncorrelated, indicating a reduction in the phonon mean free path in uncorrelated (defect) structures. Figure 4-9a shows the change in the measured thermal conductivity of silicon nanowires with their diameter, with the lowest conductivity in the narrowest nanowires as expected as the scattering of phonons at the boundaries becomes more significant. However, the temperature dependence of the conductivity is shown to change with the nanowire diameter in Figure 4-9b, from the expected Debye T^3 law scaling for 115 nm nanowires to roughly a T scaling for the smallest 22 nm nanowires. This indicates a change in the dominant mechanism for heat transport. It is proposed that this might be due to the presence of a silicon oxide layer on the surface of the wire, for which much higher phonon frequencies are supported [80]. In very narrow nanowires (about 2 nm in diameter) the thermal conductivity has been expected to actually increase due to phonon confinement effects. This increases the frequency of the longest wavelength phonons which consequently transport a greater amount of thermal energy.

Carbon nanotubes are of great interest due to their remarkable thermal properties [82]. The predicted thermal conductivities of these nanostructures are strongly dependent on their morphology, with chirality [76, 83], defects [71] and the number of walls [71] playing a vital role. Single-walled carbon nanotubes (SWCNTs) are expected to have particularly good heat conduction properties. However, the thermal

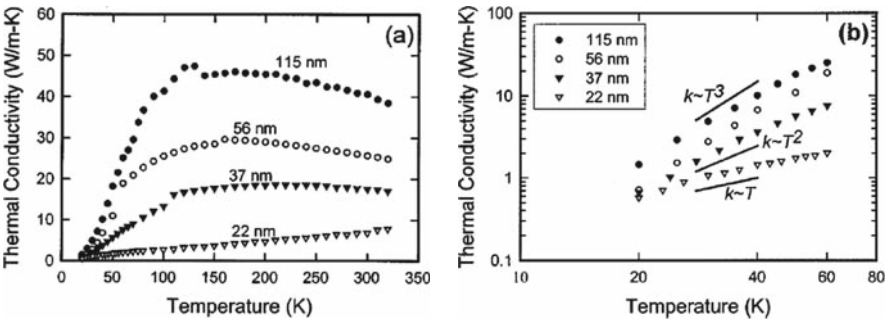


Figure 4-9. (**a**) Measured thermal conductivity of different diameter Si nanowires. The number beside each *curve* denotes the corresponding wire diameter. (**b**) Low temperature experimental data on a logarithmic scale. Also shown are T^3, T^2, and T^1 *curves* for comparison (Reprinted with permission from [Li D, Wu Y, Kim P, Shi L, Yang P, Majumdar A (2003) Thermal conductivity of individual silicon nanowires. Appl Phys Lett 83:2934]. Copyright [2003], American Institute of Physics)

conductivity of SWCNTs is not easily measured experimentally, with studies usually on bundles of nanotubes. The conductivity of bundles is much lower than single nanotubes as it is dominated by the interfacial resistance between nanotubes [70]. This inability to efficiently transfer heat between nanotubes and the surrounding medium has lead to lower than expected thermal conductivities in carbon nanotube composites. However, so high is their conductivity that a 1% weight fraction of SWCNTs embedded in an industrial epoxy can increase the thermal conductivity of the composite by 125% [84]. NEMD simulations have been used to determine

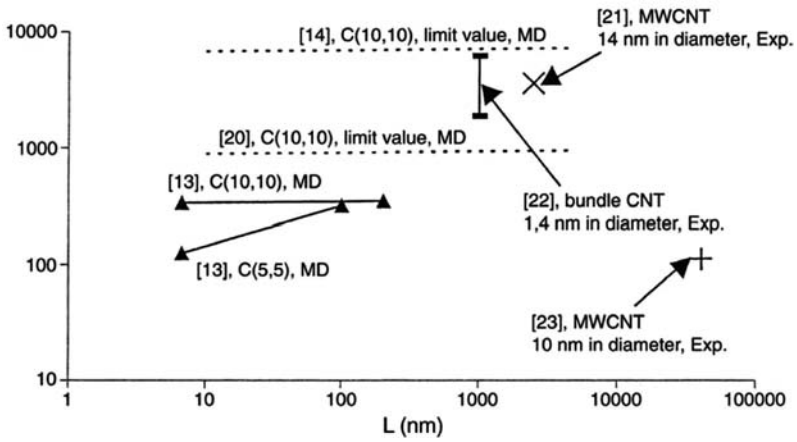


Figure 4-10. Comparison of thermal conductivity for different carbon nanotubes as a function of their length *L*. There is a wide disparity in the values obtained by different methods. The *triangular* data points are predictions from NEMD, the *horizontal dotted lines* are predictions from equilibrium MD and the crosses are experimental results. (Reprinted from [Chantrenne P, Barrat JL (2004) Analytical model for the thermal conductivity of nanostructures. Superlattice and Microstructures 35:173] with permission from Elsevier)

the thermal conductivity of SWCNTs [70, 72–75] and multi-walled carbon nanotubes (MWCNTs) but the predictions range from several hundred to 6600 W m^{-1} K^{-1} [71]. This typically wide discrepancy in results is illustrated in Figure 4-10. Lukes and Zhong [85] primarily assign this confusion to length effects. However, the importance of quantum effects is not clear [86, 87] as it is not entirely apparent what the Debye temperature is for SWCNTs. Although the Debye temperature for diamond is 2230 K (see Table 4-1) it is argued that the Debye temperature may be closer to that of graphite (reported as 420 K in [87]) although values as low as 30 K for nanotubes have been suggested [87]. Lukes and Zhong [85] report that quantum corrections to NEMD results yield qualitative agreement with experimental results. Overall, the preponderance of NEMD simulations are concerned with the study of carbon and silicon.

4.4. ISOTHERMAL CONCURRENT MULTISCALE METHODS

The number of degrees-of-freedom and femtosecond time step in MD simulation place a heavy constraint on their maximum length and time scale. There is always a compromise between accurately representing a real system and computational practicality, even in the case of nanostructures. The time and length scales are inherent restrictions of atomistic models, but this is not the case for statistical or continuum models, which represent a system by the evolution of average quantities and their distributions. The advantage of such higher level models is that systems of realistic size can be modeled over realistic time scales. However, in many problems the length and time scales are strongly coupled with the details of the atomistic processes (e.g. the dislocation creep properties of materials are determined by the relatively long-range interaction between dislocations in the crystal lattice). These details are increasingly important in areas where the material response is highly non-linear (e.g. near cracks tips and other stress concentrators, around atomic scale defects such as dislocations, grain boundaries and other interfaces) and/or highly constrained (e.g. thermal transport in nanostructures). Therefore it is often desirable to retain atomistic detail in these areas but represent the material response in the linear regime using a coarse-grained (CG) description (e.g. Fourier's law for heat transport). These models are generally described as concurrent multiscale models, as the different material models are strongly coupled and evolve and interact in parallel.

This approach has been the subject of a significant body of work over the last two decades. The most widely adopted of these methodologies has been the quasicontinuum method [9], which embeds a molecular mechanics model within a finite element model of the far field. The atomistic model is described as molecular mechanics (as opposed to MD) as the methodology is developed for zero-temperature simulation. Therefore the only degrees-of-freedom in the system are the positions of the atoms. In the finite element region the atomic positions are interpolated from the positions of the nodes using standard finite element shape functions. The two models are coupled by the requirement that atoms on the atomistic/continuum interface occupy the same relative positions as their associated

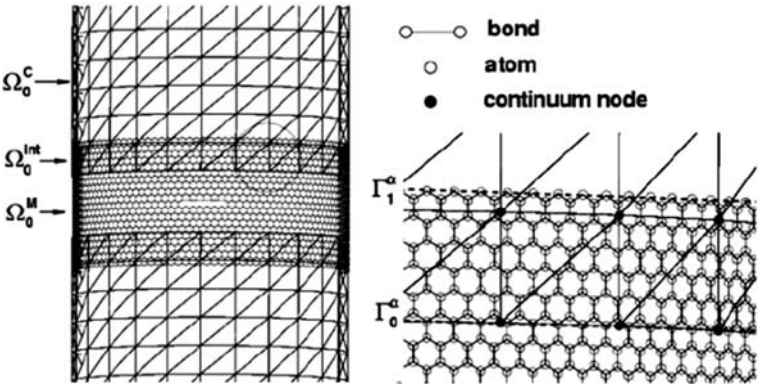


Figure 4-11. Bridging domain model for a nanotube; finite elements are indicated by *lines* that connect continuum nodes (Reprinted from [Xiao SP, Belytschko T (2004) A bridging domain method for coupling continua with MD. Comput Methods Appl Mech Engrg 193: 1645] with permission from Elsevier)

nodes in the finite element model (see Figure 4-11). Some complications arise due to the introduction of "ghost forces" at the interface due to the truncation of the interatomic potential but these can be compensated for [8].

A number of methods have been proposed to extend this methodology (and others) to finite temperature. There are some examples of this approach in the modeling of fluids [89–91] but in general the issues for fluids are not relevant to the study of solids. Abraham et al. [4] have considered the concurrent coupling of quantum, atomistic, and continuum scales in the fracture of silicon. The zero-temperature quasicontinuum model has also recently been extended to include electronic degrees-of-freedom for the simulation of metals [92]. In addition, coupling methods for quantum and atomistic models of heat transport have been proposed for non-isothermal modeling at low temperatures [39]. However, here it is assumed that the atomistic model is purely classical in nature in all respects and that the fundamental material response is given by the MD simulation from which any higher level model must be derived.

The primary issues that need to be addressed in moving from zero-temperature to finite temperature multiscale simulation are:

1. the *dynamics* of the CG nodes. The thermal vibrations of the missing degrees-of-freedom in the CG model need to be accounted for. This temperature dependence must be incorporated into the CG inter-node potential which defines the dynamics of the nodes. The derivation of this potential is discussed in Section 4.4.1.
2. *consistency* between the thermodynamic properties of the continuum and the atomistic model, e.g. the thermal conductivity and thermal expansion coefficient in the continuum model are identical to those produced in the MD model. Some of these properties can be derived from the CG potential (see Section 4.4.2). As discussed in Section 4.3.3, the thermal conductivity is dependent on the size of the atomistic region. It is expected that the atomistic region will typically be smaller than the phonon mean free path, and therefore the thermal conductivity

is not that of a bulk sample. For ballistic models, the continuum must be compatible with the atomistic region, so unfortunately this reduced finite-size thermal conductivity must be adopted throughout the simulation.

3. *smooth information transfer* across the atomistic/continuum interface. Ideally the propagation of phonons from the atomistic to the continuum region (and vice versa) would be seamless. However high frequency atomic vibrations that exist in the atomistic model cannot exist in the higher level region due to the absence of the finer length scales there. This results in the reflection of high frequency phonons from the boundary and consequently there is a net flow of heat from the continuum to the atomistic region leading to localized heating. This must be prevented, but it must be recognized that there will always be a loss of information when moving from a fine-scale to a coarse-scale model as this is an inherent feature of multiscale models. This is discussed in Section 4.4.3.

Note that these are not important issues for metal models (see Sections 4.5.3 and 4.5.4) which impose the thermal conductivity on the simulation.

The models discussed in this section are referred to as near-equilibrium models as the flow of heat between the atomistic and continuum regions is not explicity accounted for, such that the concurrent simulation as a whole retains a constant average temperature or the continuum region is isothermal. When conducting multiscale simulations it is important to decide what the function of the CG region is. Two different classes of functionality are identified here: dynamics and sampling. A dynamic CG region is inertial, e.g the finite element nodes vibrate thermally and transport (lower frequency) lattice vibrations. This is important if the aim of the simulation is to determine information about the dynamic properties of a structure where the boundaries of the CG region are "seen" within the simulation time (e.g. resonance of NEMS [12]) or to allow the transmission of dynamic information between atomistic regions (e.g. two cracks propagating simultaneously). Otherwise the main function of the CG region is to sample the average response of an ensemble of atoms and provide a representative environment for the evolution of the atomistic region. In this case the CG region does not need to retain detailed inertial information about individual lattice waves and the magnitude of thermal vibrations of some or all of the atoms can be represented by their standard continuum measure, temperature. It will be argued that this latter condition is essential for the nonequilibrium simulations presented in Section 4.5. However, here we consider the development of purely CG dynamics in the context of simulations that are close to thermal equilibrium.

4.4.1. Coarse-Grained Dynamics

The first issue to address in dynamic CG models is: what are the dynamics of the CG nodes? This has been addressed by a number of authors [10, 12, 93–97]. There are a number of routes to the same result. Firstly, we assume that the partition function of the system is unaffected by the coarse-graining process. The classical dynamics of a set of N atoms can be defined by its Hamiltonian (kinetic plus potential energy)

$$H(\{\underline{x}_i\},\{\underline{p}_i\}) = \sum_{i=1}^{N} \frac{|\underline{p}_i|^2}{2m} + V(\{\underline{x}_i\}) \tag{4-36}$$

where $\underline{p}_i = m\underline{\dot{x}}_i$ is the momentum of particle i. The continuous classical partition function for this Hamiltonian is defined as [94]

$$\begin{aligned} Z &= \tfrac{1}{h^{3N}} \int_{-\infty-}^{\infty} \int_{\infty}^{\infty} e^{-\beta H(\{\underline{x}_i\},\{\underline{p}_i\})} d\{\underline{x}_i\} d\{p_i\} \\ &= \frac{1}{\Lambda^{3N}} \int_{-\infty-}^{\infty} e^{-\beta V(\{\underline{x}_i\})} d\{\underline{x}_i\} \end{aligned} \tag{4-37}$$

where the momentum degrees-of-freedom can be integrated out exactly and $\Lambda = \sqrt{\frac{2\pi\beta}{m}}\hbar$ is the de Broglie wavelength.

Now consider a CG system where the degrees-of-freedom of a subset of n_s *slave* atoms $\{\underline{x}_s\}$ are not independent but determined by interpolation from a reduced subset of n_r *representative* atoms $\{\underline{x}_r\}$, where necessarily $n_s + n_r = N$. The CG Hamiltonian is consequently given by

$$H_{\mathrm{CG}}(\{\underline{x}_r\},\{\underline{p}_r\}) = \sum_{r=1}^{n_r} \frac{|\underline{p}_r|^2}{2m_{\mathrm{CG}}} + V_{CG}(\{\underline{x}_r\}) \tag{4-38}$$

where m_{CG} is the CG mass of the representative atoms (assuming uniform coarse-graining) and $V_{\mathrm{CG}}(\{\underline{x}_r\})$ is the CG potential. The associated CG partition function can be expressed in a similar form to (4-37) such that

$$Z = \frac{1}{\Lambda_{\mathrm{CG}}^{3n_r}} \int_{-\infty}^{\infty} e^{-\beta V_{CG}(\{\underline{x}_r\})} d\{\underline{x}_r\} \tag{4-39}$$

where $\Lambda_{\mathrm{CG}} = \sqrt{\frac{2\pi\beta}{m_{\mathrm{CG}}}}\hbar$. The CG potential can be determined from the equivalence of (4-37) and (4-39) in theory but this is difficult in practice. To make further progress we assume that the interatomic potential is harmonic such that $\phi(r) = {}^1/_2\alpha(r-a)^2$ where α is the spring constant and a is the lattice spacing. Consider three atoms in a one-dimensional chain with positions $[x_1, x_2, x_3]$. Coarse-grain the system by taking the representative atoms to be $\{x_r\} = [x_1, x_3]$ and enslave the central atom $\{x_s\} = [x_2]$. From (4-37) therefore

$$
\begin{aligned}
Z &= \frac{1}{\Lambda^3}\int_{-\infty}^{\infty} e^{-\beta V(\{x_i\})} d\{x_i\} \\
&= \frac{1}{\Lambda^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\beta\left({}^1/_{2\alpha(x_3-x_2-a)^2}+{}^1/_{2\alpha(x_2-x_1-a)^2}\right)} dx_1 dx_2 dx_3 \\
&= \frac{1}{\Lambda^2(\beta\hbar\omega_0)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-{}^1/_{4\beta\alpha(x_3-x_1-2a)^2}} dx_1 dx_3 \\
&= \frac{1}{\Lambda_{\mathrm{CG}}^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\beta V_{\mathrm{CG}}(x_1,x_3)} dx_1 dx_3
\end{aligned}
\tag{4-40}
$$

where the positional degree-of-freedom of the slave atom, x_2, has been integrated out exactly and the fundamental frequency of vibration of the system $\omega_0 = \sqrt{\frac{2\alpha}{m}}$ has been introduced. The last two lines of (4-40) give an expression for the CG potential

$$
\begin{aligned}
V_{\mathrm{CG}}(x_1,x_3) &= -\frac{1}{\beta}\ln\left[\frac{\Lambda_{\mathrm{CG}}^2}{\Lambda^2}\left(\frac{k_BT}{\hbar\omega_0}\right)e^{-{}^1/_{4\beta\alpha(x_3-x_1-2a)^2}}\right] \\
&= {}^1/_{4\alpha}(x_3-x_1-2a)^2 + k_BT\ln\left(\frac{\hbar\omega_0}{k_BT}\right) + 2k_BT\ln\left(\sqrt{\frac{m_{\mathrm{CG}}}{m}}\right) \\
&= (n_s+1)\phi(\bar{r}) + n_sk_BT\ln\left(\frac{\hbar\omega_0}{k_BT}\right) + {}^1/_2 n_rk_BT\ln\left(\frac{m_{\mathrm{CG}}}{m}\right)
\end{aligned}
\tag{4-41}
$$

where the expressions have been generalized in the final line for any number of slave and representative atoms, and $\bar{r}$ ($= {}^1/_2(x_3 - x_1)$ in this case) is the average spacing between slave atoms. Assuming that the first line of (4-41) can be extended to a general anharmonic potential this gives

$$
V_{\mathrm{CG}}(\{\underline{x}_r\}) = -\frac{1}{\beta}\ln\left[\left(\frac{\Lambda_{\mathrm{CG}}^{n_r}}{\Lambda^N}\right)^d \int_{-\infty-}^{\infty} e^{-\beta V(\{\underline{x}_r\},\{\underline{x}_s\})} d\{\underline{x}_s\}\right] \tag{4-42}
$$

where d is the dimension. In this formulism the CG potential is commonly known as the potential of mean force (PMF) [98]. Although the factor $\left(\frac{\Lambda_{\mathrm{CG}}^{n_r}}{\Lambda^N}\right)^d$ has the dimensions of (length)$^{-(N-n_r)d}$ it is often omitted as it only makes a constant (although temperature-dependent) contribution to the CG potential which is of no consequence for isothermal simulation.

Generalizing the last line of (4-41) in three-dimensions gives an expression for the classical CG potential

$$V_{\mathrm{CG}}^{\mathrm{classical}}(\{\underline{x}_r\},T) = V(\{\underline{x}_r\}) + k_B T \sum_{r=1}^{n_s} \sum_{q=1}^{3} \ln\left(\frac{\hbar\omega_{rq}}{k_B T}\right) + {}^1/_2 n_r k_B T \ln\left(\frac{m_{\mathrm{CG}}}{m}\right) \tag{4-43}$$

where ω_{rq} is the qth eigenfrequency determined for the rth slave atom from (4-26) and $V(\{\underline{x}_r\})$ is calculated from the positions of the representative atoms by assuming the slave atoms are uniformly spaced between them. The first term in the CG potential (4-43) is the zero-temperature potential energy of the body derived for the zero-temperature quasicontinuum method [9]. The second term accounts for the contribution to the free energy from the thermal vibrations of the missing slave atoms.

The last (mass) term in (4-43) is often neglected as it is constant for isothermal simulations. It is clear that the partition function places no particular restriction on the choice of the CG mass. One approach is to assume that $m_{\mathrm{CG}} = m$, in which case the last term disappears even for non-isothermal conditions. However, this choice will mean that the speed of sound in the material is a function of the coarse-graining. Another approach is the lumped mass model where it is assumed that $n_r m_{\mathrm{CG}} = Nm$ to conserve mass. This ensures that the wave speed is independent of the coarse-graining. However, the above calculations assume that the coarse-graining is uniform, i.e. the density of representative atoms in the CG region is constant. Dupuy et al. [96] have considered the non-uniform case in more detail.

The dynamics of the representative atoms evolve in the usual way

$$m_{\mathrm{CG}} \ddot{\underline{x}}_r = -\frac{\partial V_{\mathrm{CG}}}{\partial \underline{x}_r} \tag{4-44}$$

which is identical to (4-1) in the fully atomistic limit ($n_s = 0$) as one would expect. However, it should be noted that the representative atoms only truly represent the time-averaged position of an atom (or group of atoms); their oscillation frequency and amplitude are not those of a normal atom.

The second term in (4-43) is dependent on the system temperature, the level of coarse-graining and, through (4-26), the local deformation of the body. This term can also be derived from the quantum partition function using the local harmonic approximation of the Einstein model for a solid [93]. This was discussed in Section 4.2.4 and is advantageous over the Debye model in that the thermodynamics are analytically tractable and only depend on the fundamental frequencies of the lattice. For a simple quantum harmonic oscillator the partition function is given by (4-9) and (4-22) to be

$$Z_n = \frac{e^{-\beta\hbar\omega_0/2}}{1 - e^{-\beta\hbar\omega_0}} = \frac{1}{2\sinh(\beta\hbar\omega_0/2)} \tag{4-45}$$

The Helmholtz free energy, $F = U - ST$, gives the available work under the constraints of constant temperature and volume, where S is entropy and the internal energy is given by (4-23), and hence is the characteristic function for the canonical

ensemble. For a single quantum harmonic oscillator in one-dimension the thermal contribution is [93]

$$F = -k_B T \ln Z_n = k_B T \ln\left[2\sinh\left(\beta\hbar\omega_0/2\right)\right]. \tag{4-46}$$

using (4-45). It is clear from the above expression that assuming that the Helmholtz free energy is unchanged by the coarse-graining process is equivalent to assuming that the (quantum) partition function remains unchanged. Therefore the difference between the free energies due to the loss of the dynamic thermal contribution from the slave atoms for the quantum CG potential is

$$V_{\mathrm{CG}}^{\mathrm{Einstein}}(\{\underline{x}_r\},T) = V(\{\underline{x}_r\}) + k_B T \sum_{r=1}^{n_s}\sum_{q=1}^{3} \ln\left(2\sinh\left(\frac{\hbar\omega_{rq}}{2k_B T}\right)\right). \tag{4-47}$$

This is identical to (4-43) in the classical limit $x = \dfrac{\hbar\omega}{k_B T} < 1$ for which $2\sinh(x/2) \approx x$. The final mass-dependent term in (4-43) is absent from (4-47) as the momentum of the atoms is not included in the phonon model.

4.4.2. Coarse-Grained Thermal Properties

It is important that the CG potential reproduces the correct thermophysical properties for consistency with the atomistic model. The most obvious contribution from thermal vibrations in the lattice is the expansion of the crystal with an increase in the temperature. The CG potential allows for an estimate of this contribution. For a one-dimensional solid at constant volume the macroscopic stress that develops in the body due to a temperature rise of ΔT is given by $\sigma_T = \alpha_T E \Delta T$ where α_T is the coefficient of thermal expansion and E is Young's modulus. This stress is equivalent to an atomic thermal expansion force of

$$f_T = -\frac{1}{n_s}\frac{\partial V_{\mathrm{CG}}}{\partial r}\bigg|_{\bar{r}=a} = -\left[\left(1+\frac{1}{n_s}\right)\frac{\partial\phi}{\partial r} + \frac{\hbar}{2}\frac{\partial\omega_0}{\partial r}\coth\left(\beta\hbar\omega_0/2\right)\right]_{\bar{r}=a} \tag{4-48}$$

where the simple one-dimensional CG potential in the final line of (4-41) has been adopted without the assumption of classical mechanics. Given that $\omega_0^2 = \dfrac{2\alpha}{m} = \dfrac{2}{m}\dfrac{\partial^2\phi}{\partial^2 r}$ and taking the coarse-graining to the macroscale ($n_s \to \infty$) one can determine the thermal expansion force. In the classical limit (4-48) gives

$$f_T^{\mathrm{MD}}(a) = -{}^1\!/_2 k_B T \frac{\phi_{,rrr}(a)}{\phi_{,rr}(a)}. \tag{4-49}$$

where $\dfrac{\partial\phi}{\partial r}\bigg|_{\bar{r}=a} = \phi_{,r}(a) = 0$ at the equilibrium zero-temperature spacing. From this expression it is immediately clear that there will be no thermal expansion in the case of a harmonic potential as $\phi_{,rrr} = 0$. For the Lennard-Jones potential

$\phi(r) = \varepsilon\left(\left(\frac{a}{r}\right)^{12} - \left(2\frac{a}{r}\right)^{6}\right)$ this yields $f_T^{\mathrm{MD}} = \frac{21k_B}{2a}T$ which predicts that the force is linearly proportional to the temperature in agreement with the macroscopic stress. Roughly $f_T^{\mathrm{MD}} \approx -\sigma_T a^2$ which suggests that $\alpha_T E = {}^{21}/_{3}k_B a^{-3}$.

This classical prediction and the full quantum prediction from (4-48) are compared with the results of a one-dimensional MD simulation in Figure 4-12a. It is expected that the simulation results should correspond with the classical CG model and this is found to be the case. The agreement is excellent at low temperatures, with only a small deviation at high temperatures. The quantum model converges to the classical model at high temperatures but is significantly different below room temperature. The quantum thermal expansion force is predicted to be finite at zero temperature, i.e. a is not the equilibrium lattice spacing at zero temperature. This is due to the existence of the zero-point energy in a quantum crystal introduced in (4-22). However, the Einstein model has been shown to be inaccurate at very low temperatures (see Figure 4-2). The predictions for a two-dimensional nickel crystal [96] are shown in Figure 4-12b. Here the crystal is allowed to expand and the change in lattice spacing with temperature recorded. The agreement for the lattice parameter is good but not as good as the constrained thermal expansion force in Figure 4-12b. This is because different values of the anharmonic potential are sampled for the case of free expansion.

It should be noted that the thermal force $f_T^{\mathrm{MD}}(r) \to \infty$ when $\phi_{,rr}(r) = 0$ [95]. This situation occurs at the inflexion (point of zero curvature) in the potential where the stiffness vanishes. For the Lennard-Jones potential this occurs at a radius of $1.109a$.

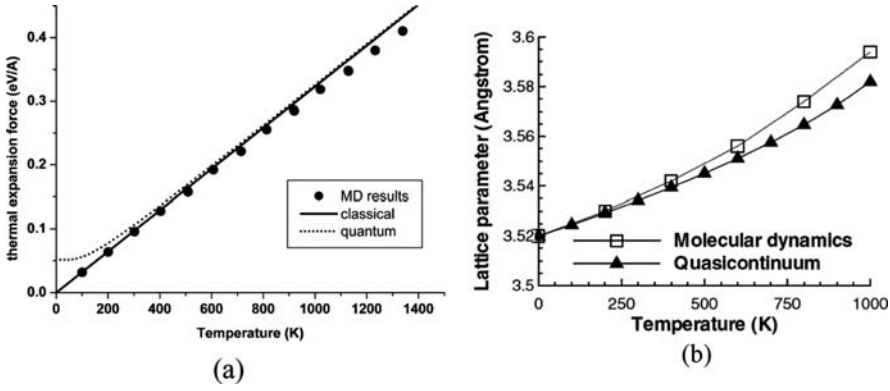


Figure 4-12. (**a**) Average thermal expansion force f_T in a one-dimensional MD simulation of Lennard-Jones atoms with $\varepsilon = 0.6$ eV, $a = 2.8$ Å and $m = 10^{-25}$ kg [95]. The *lines* are the predictions from the classical (4-49) and quantum (4-48) CG potentials. As expected, the quantum and classical predictions are indistinguishable at high temperatures. Note that the quantum predictions do not go to zero at 0 K due to the existence of a finite zero point energy (see 4-22). (**b**) Lattice parameter of nickel as a function of temperature using an embedded atom model (EAM) potential. The melting temperature of this potential is 1478 K. The CG (quasicontinuum) calculation involved a cell of dimensions 200 × 100 nm with a regular mesh containing 50 nodes (Reprinted (Fig 1, pg 115404-1) with permission from [Dupuy LM, Tadmor EB, Miller RE, Phillips R, Phys Rev Lett, 95,2005. Copyright by the American Physical Society. http://link.aps.org/abstract/PRL/v95/p060202)

There is always a finite probability that this state will be sampled during the course of a simulation. At low temperatures the chance is small with the possibility rapidly increasing with temperature. This problem arises from the nature of the quadratic approximation in the determination of the CG potential. If the full nonlinear potential is calculated from (4-42) then this situation does not arise, although this is not a practicable solution for CG simulation.

4.4.3. Boundary Conditions for the Atomistic/Continuum Interface

One of the critical issues in finite-temperature multiscale simulation is the effect of interfaces between different length scales on phonon heat transport. The requirements of the phonon boundary condition depend on the nature of the problem. For dynamic coarse-graining, lower frequency (coarse-scale) phonons must be transmitted into the continuum, whereas the higher frequency (fine-scale) phonons that are not supported in the continuum should be absorbed and generated according to the canonical ensemble. The boundary conditions for phonons for a quasi-static continuum are simpler in that all incident phonons simply need to be absorbed and generated. Two main classes of approach are identified here:

(i) Memory kernel methods. Impedance boundary conditions are based on the generalized Langevin equation (GLE) [30, 99] which are an exact derivation for harmonic interatomic interactions. For an atom i in the interfacial region this is typically expressed as [99]

$$m\ddot{\underline{x}}_i = -\frac{\partial V}{\partial \underline{x}_i} + \sum_j \int_0^t \beta_{ij}(t-\tau)\dot{\underline{x}}_i(\tau)d\tau + \sum_j \beta_{ij}(t)\underline{x}_j(0) + \underline{R}(t) \tag{4-50}$$

where $j \neq i$ represents the other atoms in the interfacial region, $\beta_{ij}(t)$ is the memory kernel function representing the history of interaction between atoms i and j, and $\underline{R}(t)$ is a random noise representing the thermal vibrations of the heat bath. The existence of the convolution integral term is due to the importance of long-time correlations in the motion of atoms in a crystal. Exact determination of the convolution integral is expensive and it is characteristically truncated over space and time usually without introducing significant errors, at least in the harmonic limit. This methodology has principally been applied to the study of "zero-temperature dynamics" in which the atoms initially have no kinetic energy and are subject to a dynamic perturbation [14, 99–102]. One advantage of "zero-temperature dynamics" is that the propagation of phonons through the medium can be easily resolved without the presence of thermal noise. The effectiveness of these impedance boundary conditions is nicely illustrated in such a case in Figure 4-13. A Gaussian-type wave perturbation of high and low frequencies is introduced at the centre of the atomistic region. The wave propagates outwards and is expected to be smoothly transmitted into the surrounding continuum. This is the case where memory kernels are used in Figure 4-13a. The reflection of the high frequency waves can be clearly seen in Figure 4-13b where there is no special treatment of the boundary conditions. One

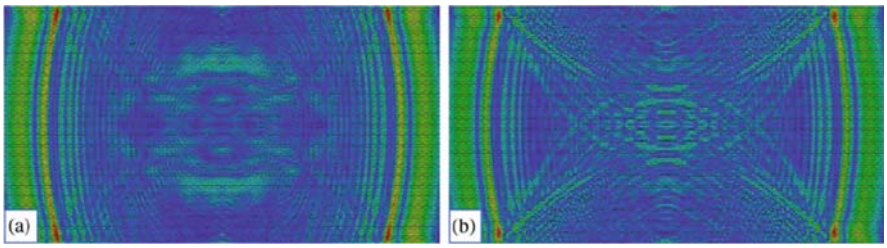


Figure 4-13. Wave propagation through the atomistic domain in the FCC lattice structure: (**a**) impedance boundary conditions are involved at the MD/continuum interface, (**b**) continuity boundary conditions (Reprinted from [Liu WK, Karpov EG, Zhang S, Park HS (2004) An introduction to computational nanomechanics and materials. Comput Methods Appl Mech Engrg 193: 1529] with permission from Elsevier)

disadvantage of memory kernels is that it is difficult to move the position of the atomistic/continuum interface during the simulation due to their dependence on the simulation history.

Methods for extending the GLE to finite temperature have been recently proposed to thermalize the system as well as absorb phonons [35, 103]. As mentioned in Section 4.3.2, this type of boundary condition holds the best opportunity for minimizing decorrelations in atomic motions at the interface due to thermostatting and avoiding the artificial Kapitza effect. However, no studies of non-isothermal systems using the GLE approach appear to have been conducted at this time. This may be complicated by the fact that the memory kernels depend on temperature [35] and therefore the local temperature history also. They are also highly potential specific and have a complicated implementation structure [35, 103].

(ii) Diffuse interface methods. A simpler method for minimizing phonon reflections is to make the transition from one length scale to another as smooth (or diffuse) as possible. This can be achieved to some extent by subtle mesh gradation, although such meshes can be complicated to generate and can add complexity to the coarse-graining algorithm. The bridging domain method [88] minimizes phonon reflection by a smooth transition from the atomistic representation to the continuum, such that the interfacial Hamiltonian is given by

$$H = \alpha H^M + (1 - \alpha) H^C \tag{4-51}$$

where H^M and H^C are the Hamiltonian's in the molecular and continuum regions respectively and α varies continuously through the interface from 1 in the purely atomistic region to 0 in the purely continuum region. This means that the two regions must overlap in the transition zone, as shown in Figure 4-11. This has the advantage of being simple to implement given the continuum dynamics, and has been demonstrated to be effective for "zero-temperature dynamics".

A diffuse phonon-absorbing interface known as *stadium damping* has been proposed by Holian and Ravelo [3]. This had been employed in an isothermal concurrent atomistic/continuum simulation by Qu et al. [104, 105]. In this approach, an atomistic region is embedded within a quasi-static finite element mesh, as shown in

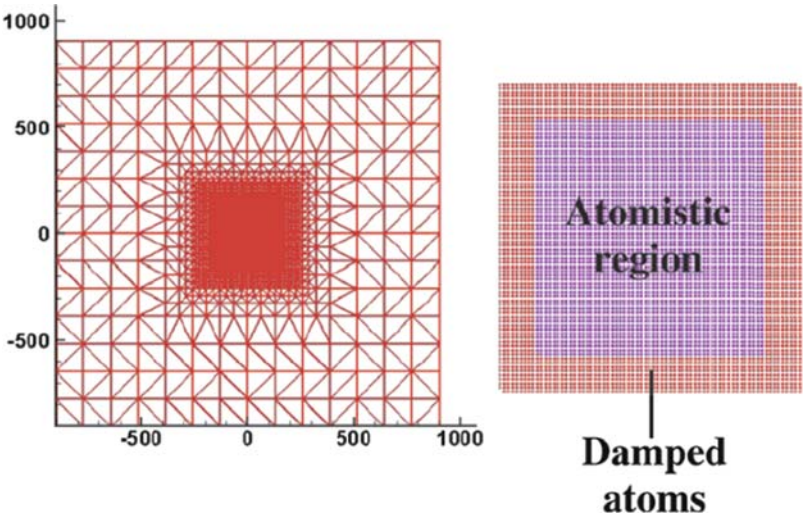


Figure 4-14. Schematic of the finite-temperature dynamic coupled atomistic/continuum simulation cell, showing the outer continuum finite element region (*left*), the damped stadium region of atoms near the atom/continuum boundary (*red*) and an undamped atomistic region in the interior (*blue*) (Reprinted from [Qu S, Shastry V, Curtin WA, Miller RE (2005) A finite temperature dynamic coupled atomisitic/discrete dislocation method. Model Sim Mater Sci Engng 13:1101]with permission from Institute of Physics)

Figure 4-14. The mesh can deform (elastically) but the nodes are not dynamic. The atoms at the outer boundary of the (red) atomistic region are effectively quasi-static in nature and deform with the mesh for continuity. The atoms in the central (blue) region are fully dynamic and unthermostatted. The quasi-static and fully dynamic atoms are separated by a (red) interfacial region of stadium damped atoms. The dynamics of these atoms are regulated by Langevin thermostats, as in (4-5), except that the strength of the damping, $\gamma(\underline{x})$, varies linearly from zero in the fully-dynamic (blue) region to its maximum value $\gamma_{\max}$ at the outermost (quasi-static) boundary. Unlike a sharp interface, the resulting diffuse interface allows phonons to penetrate into it, whereupon they are gradually thermostatted out of the system. Stadium damping is advantageous in that it is simple to implement and it is independent of the interatomic potential. It also thermalises the boundary (and hence the fully-dynamic region) to a specified temperature and has been shown to generate the canonical ensemble in isothermal simulation [104]. Stadium damping also allows the outer atomistic boundary to be fixed/quasi-static which is essential for embedding in a quasi-static CG region. The effectiveness of stadium damping to absorb phonons is illustrated in Figure 4-15.

(iii) alternative methods. Other treatments for the boundary conditions do not fall into the previous categories. Gill et al. [95] proposed a rapid thermalisation method, although this requires a small time step and is computationally intensive. The remaining methods have been shown to be effective for "zero-temperature dynamics". Tang [106] has proposed computationally efficient velocity interfacial conditions in which a high-order Taylor expansion is used to predict the atomic motions at the interface [106]. It is claimed that this method has a higher performance than memory kernel methods for strongly nonlinear problems and has the

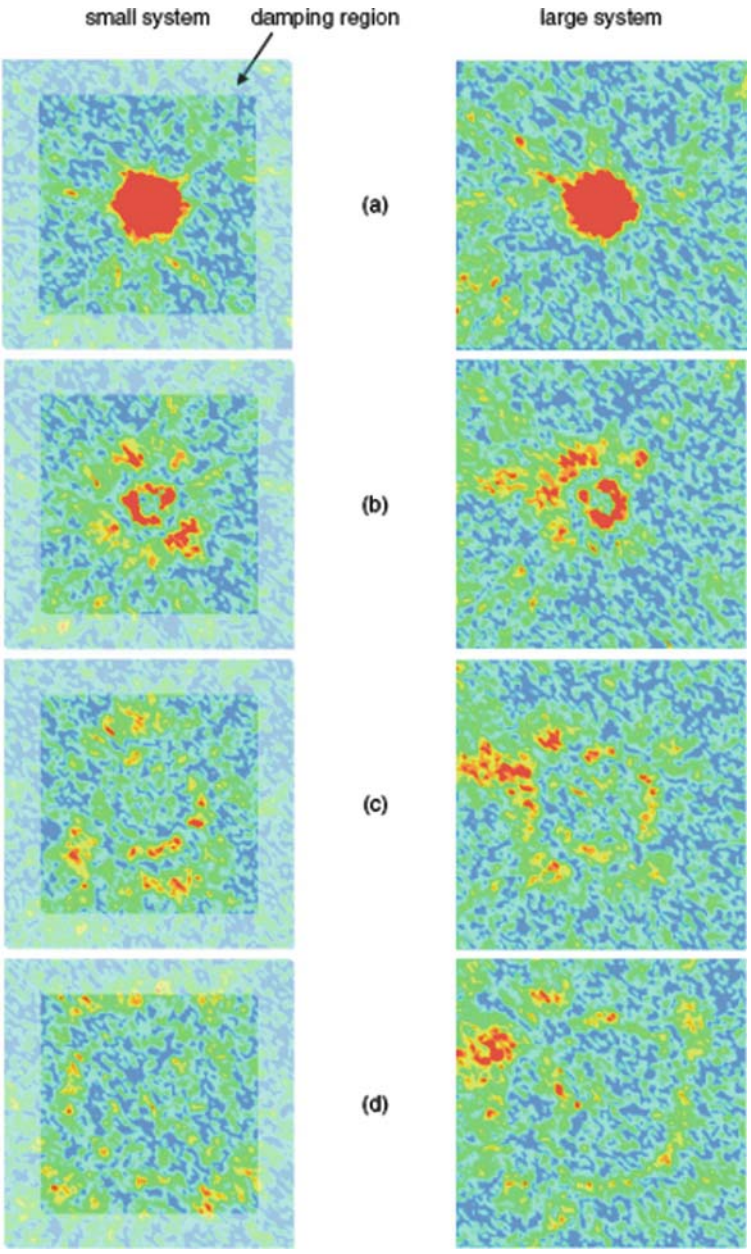


Figure 4-15. A high energy pulse is introduced at the centre of a stadium damped atomistic simulation (*left side*) and a much larger (reference) atomistic simulation (*right side*) at 100 K. The *snapshots* are shown at different times (**a**) $t = 0$, (**b**) $t = 14$, (**c**) $t = 28$, (**d**) $t = 42$. The phonons are absorbed by the stadium damping region (*left*) or propagate out of view (*right*). The differences are due to thermal fluctuations. The *color scale* indicates the magnitude of the displacements (Reprinted from [Qu S, Shastry V, Curtin WA, Miller RE (2005) A finite temperature dynamic coupled atomisitic/discrete dislocation method. Model Sim Mater Sci Engng 13:1101]with permission from Institute of Physics)

advantage of being local in time and space. Li et al. [107] use a Perfectly Matched Layer (PML) which adds a linear damping term and change in stiffness term to the dynamics of the interfacial nodes. They investigate a harmonic potential but present derivations for higher order potentials. The method is extended to finite temperature [108] but the precise efficacy of the boundary conditions is not investigated. Namilae et al. [109] have proposed and analyzed a number of simple absorbing conditions based on the non-linear wave equation [109] and implemented them in a concurrent multiscale model.

4.4.4. Isothermal Dynamic Multiscale Models

Dupuy et al. [96] utilize the CG potential (4-43) (the third term is neglected at constant temperature) to simulate the isothermal nanoindentation of nickel at different temperatures using a non-uniform coarse-graining. All the atoms and CG finite element nodes vibrate with thermal energy. No special boundary conditions at the atomistic/continuum interface are introduced to tackle the issue of phonon reflection. This is addressed by the application of a single global Nosé-Hoover thermostat to the dynamics of all the degrees-of-freedom (i.e. nodes and atoms). This prevents the atomistic region heating up due to the confinement of high frequency phonons. Strictly the natural increase in thermal energy of the simulation (due to applied loads) should not be removed by a thermostat, but allowed to be transported to the boundaries. However, given the subject material is a metal, the thermal transport processes inherent in the MD simulation would drastically underestimate those in the true metal anyway, due to the absence of conducting electrons in the classical model. (This problem is addressed in Sections 4.5.3 and 4.5.4). Assuming the thermal equilibration processes are rapid, the isothermal assumption is reasonable. The performance of the Nosé-Hoover thermostat under these conditions could be questionable given that the CG simulation will contain multiple inherent time scales. Advanced thermostats such as the recursive multiple thermostats [32, 95] may be more pertinent to this type of problem. Additionally, as discussed in Section 4.2.2, a global thermostat will not ensure a uniform temperature in a nonequilibrium simulation such as this. It will only drive the system towards a given average temperature. In doing this it is possible that some regions are cooled below the average temperature to offset the increase in temperature of other regions.

Heat transport in dynamic CG models has been studied by Curtarolo and Ceder [97], using a CG potential similar to (4-43) (again without the final mass term) using a lumped mass approximation. They find excellent correspondence between fully atomistic and CG simulation predictions for the elastic properties of the system. The thermal conductivity in CG systems is also considered. It is found that this is always underestimated in the CG region, as one would expect with the reduction in the available phonon modes for heat transport. These calculations are in the steady state where the specific heat capacity of the CG nodes is of no consequence. To fully model heat transport in a dynamic CG system is impossible as the specific heat capacity of each degree-of-freedom is always k_B in the classical MD model (see Section 4.2.5). Therefore the heat capacity of a dynamic CG system will not be constant but vary in space with the degree-of-freedom density (level of coarse-graining).

Rudd and Broughton did a lot of the early work on coarse-graining a harmonic Hamiltonian [10] at zero-temperature, and have recently extended this with an anharmonic correction to incorporate nonlinear effects at finite temperature [12]. They investigate the effect of coarse-graining on the phonon dispersion relationship and phonon reflection at interfaces (see Figure 4-1b). They show that the distribution of mass to the CG nodes can have a significant effect on the response of the system. They distribute the mass through the finite element shape functions which seems preferable to the lumped mass approach which exacerbates the reflection of phonons at CG interfaces. In this case it is necessary to retain as much information about the dynamics of the system as possible. However, in general, retaining the dynamic motions of CG nodes is computationally expensive and need only be considered if it is a necessary requirement of the CG region.

Other dynamic approaches which do not use a CG potential have also been proposed [88], the most widely applied of which is the bridging scale method [14, 100, 101]. In this method, a coarse-scale finite element continuum exists everywhere such that it overlaps the atomistic region, where a fine-scale also exists. The displacement field is decomposed into coarse and fine scales in this region. The coarse-scale at all points is defined by the displacements of the finite element nodes via the shape function in the usual manner. The fine scale is the difference between the total displacement and the coarse-scale displacement, such that it represents the part of the displacement field that the coarse-scale cannot represent. The fine-scale displacement is chosen such that it minimizes the least square difference between the atomistic and coarse-scale displacements. This results in the coarse-scale and fine-scale dynamics being coupled in the atomistic region. The dynamics in the purely coarse-scale region evolve under the action of a zero-temperature potential not entirely dissimilar to $V(\{\underline{x}_r\})$ in (4-43). This is principally a "zero-temperature dynamics", in the sense that the method has primarily been used to look at the propagation of elastic waves in bodies with zero initial kinetic energy. The thermal vibrations of the missing degrees-of-freedom, represented by the temperature-dependent terms in (4-43), do not therefore need to be accounted for. The bridging scale methodology has been applied to the evolution of temperature in coupled atomistic/continuum simulations [54]. This proposes a method for projecting the kinetic energy of atoms onto a coarse-scale temperature grid. A model for the evolution of the temperature in a coarse-scale region is developed based on the discretized wave equation. This harmonic approach has the advantage of being analytically tractable but it is not applicable to finite temperature heat transport, which is purely due to anharmonic interactions. It does, however, allow for the exact solution of the phonon reflection problem at the atomistic/CG interface (see memory kernel methods in Section 4.4.3.)

4.5. NON-ISOTHERMAL CONCURRENT MULTISCALE METHODS

Heat transport and the conservation of thermal energy are requirements of non-isothermal multiscale models. The CG regions can be either dynamic or quasi-static, but it is essential that the kinetic energy of the missing degrees-of-freedom is represented in some way. In the quasi-static approach, the CG nodes have no kinetic

energy and the kinetic energy of all the atoms (representative and slave) represented by the state variable temperature. In the dynamic approach, the CG nodes (representative atoms) have kinetic energy and only the missing kinetic energy of the slave nodes is represented by the elemental temperature. The evolution of temperature within the CG region can be readily calculated from (4-31) with relatively little computational overhead given a knowledge of the thermal conductivity. This suggests two approaches: one in which the thermal conductivity of the atomistic simulation is taken to be the correct one (applicable for purely phonon heat transport such as that found in insulators), and one in which the thermal conductivity of the continuum is prescribed and is then imposed on the atomistic simulation (suitable for non-classical heat transport such as that found in metals). These are consequently described as insulator and metal models respectively. The objective of this section is to look at multiscale methods for tackling nonequilibrium, non-isothermal problems in solids.

4.5.1. Quasi-Static Phonon Models for Insulators

Although most multiscale models are dynamic, it is likely that sampling will be a sufficient requirement for most CG regions in multiscale simulations. Gill and Jolley [95] have proposed an extension to the isothermal model of Qu et al. [104, 105] (see Figures 4-14 and 4-15) to inhomogeneous thermal problems. They avoid the Kapitza problem by regulating the atomistic-continuum boundary indirectly using adjacent thermostatted zones that are not directly part of the atomistic-continuum model (see Figure 4-6). Stadium damping thermostats are used to absorb phonons at the boundary and thermalize the boundary, where use of a local thermostat such as Langevin is useful in imposing inhomogeneous thermal boundary conditions (see Section 4.2.2). For a purely atomistic NEMD simulation in Figure 4-6, the thermostatted regions (TRs) are simply used to regulate the temperature at the boundary of the true dynamics region (TDR), with any artificial Kapitza boundary effects occurring in the intermediate buffer regions (BRs). The TDR is the only region of interest for the numerical experimentalist. It is simple to couple the TDR of such a simulation to a quasi-static continuum region evolving under (4-31). Now the temperature at the boundaries of the TDR are not known. They are determined naturally during the simulation from the energy coupling between the atomistic and continuum regions. The temperatures of the TRs are regulated to ensure conservation of thermal energy such that

$$Q_q \dot{T}_L = \int_0^t (q_L - \langle q \rangle_L) dt \quad Q_q \dot{T}_R = \int_0^t (q_R - \langle q \rangle_R) dt \tag{4-52}$$

where T_L and T_R are the thermostat target temperatures (as in 34), Q_q is a constant which determines the responsiveness of the thermostats, and $q - \langle q \rangle$ is the difference between the heat fluxes in the continuum and the atomistic regions along a shared boundary at the centre of the BR. The integrals ensure that heat is conserved exactly

over time. Two examples of fully transient simulations are shown in Figure 4-16. The temperature in the atomistic TRs and BRs is not shown as these purely exist for the purposes of controlling the boundary conditions of the TDR. Figure 4-16a illustrates the thermal response of a system which is initially at a uniform 20 K and then the temperature at the lefthand continuum boundary is instantly raised to 40 K. The thermal noise in the atomistic profile is minimized by averaging over time and repeated simulations. The predictions of a purely continuum model are also shown, where the thermal conductivity is determined to be $k(T) = k_0 + \nabla k T$ with $k_0 = 1.037/\sqrt{m}$ and $\nabla k = 0.021/\sqrt{m}$W m^{-1} K^{-1} where m is the (dimensionless) atomic mass, taken in this case to be one. The two models agree well, indicating that Fourier's law (4-30) is valid for three-dimensional Lennard-Jones MD simulations even under high temperature gradients. The steady state solution is not completely linear due to the change in conductivity with temperature. Averaged over very long times the steady state profile in the coupled system is continuous with no visible local disturbance at the atomistic/continuum interface. The atomistic-continuum interfaces on the left and right-hand sides evolve under identical algorithms, showing that they are equally valid for heat flow in either direction. As Fourier's law appears to be valid, coupled simulation of homogenous atomistic regions appears to be unnecessary. Figure 4-16b illustrates the case of an inhomogeneous atomistic region which contains a layer of atoms with 10 times the mass of the others. This mass difference leads to phonon reflection at the interface and a drop in the thermal conductivity between high and low mass regions. This is a real Kapitza effect, as opposed to the artificial Kapitza effect previously encountered due to thermostats (see Figure 4-5), and is not present in continuum models. This effect has been incorporated in the continuum model here by recalibrating the thermal conductivity between cells of different mass to a much lower value with $k_0 = 0.0001$ and $\Delta k = 0.001$. The resulting continuum model again nicely reproduces the atomistic results. This is a three-dimensional atomistic simulation of a one-dimensional heat transport problem. For a fully three-dimensional heat transport problem, the issue of the anomalous thermal expansion of the redundant TR and BR regions needs to be addressed as these are typically at more extreme (high or low) temperatures than those in the TDR.

Knowledge of the heat capacity, thermal conductivity and expansitivity is usually required for such models to ensure compatibility between the atomistic and continuum regimes. It has been shown that the heat capacity is known exactly for classical dynamics (k_B per degree-of-freedom) and that the thermal expansitivity can be approximated from the interatomic potential (see Figure 4-12). However, as seen in Section 4.3, transport coefficients such as the thermal conductivity are more complex and not readily approximated from atomistic theory. Therefore it is necessary to calibrate the thermal conductivity from direct molecular simulation. This can be readily done for a particular situation. However, the conductivity is often a function of other state variables (such as strain [56]) and it is not always possible (or desirable) to determine the full dependence of a material parameter on all of these state variables. Therefore, under some conditions, it can be attractive to determine such parameters during the course of the simulation. Give the atomistic

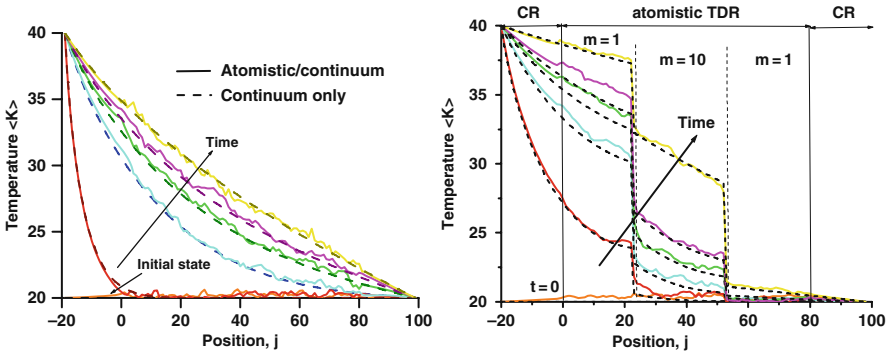


Figure 4-16. Snapshots of the temperature profile evolution for an entirely continuum model (*dashed lines*) and a coupled atomistic/continuum model, where the atomistic region (TDR) is $0 \leq < j \leq 80$ and the rest is the continuum region (CR). (a) The system is initially at a uniform 20 K but then the temperature of the left-hand (continuum) boundary is instantly changed to 40 K. The final (*yellow*) curve is the steady state temperature profile. This is curved due to the dependence of the thermal conductivity on temperature. The purely continuum model and coupled model are indistinguishable, implying Fourier's law (4.31) applies for Lennard-Jones MD under large temperature gradients. (**b**) The response of the same system except that a central layer of atoms with 10 times the mass of the other atoms has been introduced into the atomistic system. The progress of heat through the body is hindered by phonon scattering at the interface between the different mass regions. The steady state curve is again shown in *yellow* (From [42])

heat fluxand temperature gradient, on-the-fly estimates can be used to determine and refine such values concurrently. The Heterogeneous Multiscale Method (HMM) [110, 111] adopts this approach in which small MD simulations are used to inform a higher-level continuum model, and in some cases exclusively used to give atomistic detail around defects etc. Missing macroscale data (such as microscopic energy

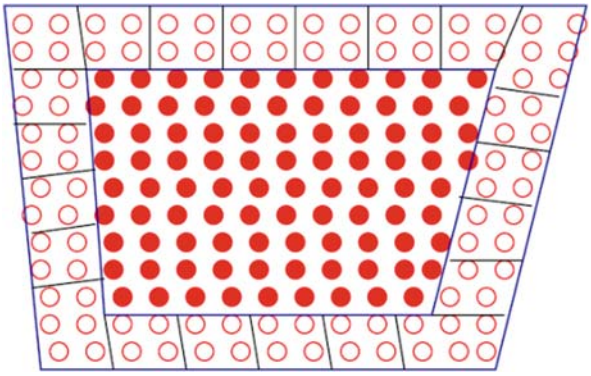


Figure 4-17. Inhomogeneous thermal boundary conditions imposed on a central atomistic system for the Heterogeneous Multiscale Method (HMM): the border region is divided into bins and a Nosé–Hoover thermostat is applied to each bin (Reprinted from [Li X, E W (2005) Multiscale modelling of the dynamics of solids at finite temperature. J Mech Phys Solids 53:1650] with permission from Elsevier)

and momentum fluxes) are estimated using an on-the-fly microscale atomistic simulation at each point where it is needed. The MD simulation is constrained so that it is consistent with the local macroscopic state. Inhomogeneous thermal boundary conditions are achieved by dividing the border region into bins and regulating the temperature of each bin using a number of Nosé-Hoover thermostats, as shown in Figure 4-17. To be practical, the size of the individual microscale simulations needs to be as small as possible, although for accurate thermal conduction the size will ideally be many times greater than the length of the phonon mean free path.

4.5.2. Dynamic Phonon Models for Insulators

Xiao and Beltyschko [88] investigated one-dimensional heat conduction using their bridging domain method. They considered the imposition of a steady state temperature gradient across the atomistic/continuum simulation. As can be seen from the interfacial condition (4-51), the dynamics of the continuum nodes is defined by a continuum Hamiltonian. This has the advantage that some of the dynamic information (about the longer wavelength vibrations) is retained, although at some computational cost. However, as mentioned previously, the kinetic energy of these nodes cannot fully represent the vibrational energy of the continuum due to the reduced number of degrees-of-freedom. The vibrational energy of the missing degrees-of-freedom is represented by assigning a temperature to each continuum cell. This temperature evolves according to the continuum energy conservation equation (4-31). The thermal conductivity is an unknown parameter and is initially determined from a fully atomistic NEMD simulation to ensure consistency between the atomistic and continuum descriptions. The transfer of energy between the atomistic and continuum regions is determined by the interfacial condition (4-51) in the overlapping region. This always removes energy from the fine-scale as the positions of the atoms are constrained to be commensurate with their associated continuum positions in this region (interpolated from the elemental shape function). This lost energy provides the energy source for the continuum temperature equation if heat flow is expected from the atomistic to the continuum region (i.e. the temperature of the atomistic region is higher than that of the continuum). Otherwise the energy is fed back into the atomistic region. The continuum temperature equation can also act as a heat source for the atomistic region. This energy is injected into the atomistic boundary using a Berendsen thermostat. The resulting temperature profiles are roughly what is expected (see Figures 25 and 27 in [88]) although the high degree of thermal noise makes the details of the correspondence hard to determine. As expected, even with a smooth atomistic-continuum interface (4-51), the Kapitza effect cannot be avoided entirely.

Fish and Li [112] have proposed a finite-temperature CG continuum model based on Generalised Mathematical Homogenization (GMH). The problem is divided into two length and time scales, a coarse-scale and a fine-scale. In the coarse-scale the body is continuous and in the fine-scale it is discrete. The displacement and interatomic force are expanded into a leading order (coarse length-scale) term and a higher order fine length-scale term. The acceleration is then expanded into terms that

include coarse and fine time-scales and length-scales. The resulting MD equations of motion yield the coupled dynamics of the fine and coarse scales. The fine-scale motion only functions at the fast time-scale. The coarse-scale motion includes the fast time-scale and the slow time-scale. The two different time-scales are separated out to give a coarse-scale wave equation and a coarse-scale thermal equation. The heat flux depends on the fine-scale and coarse-scale velocities. The coarse-scale is represented by a dynamic finite element mesh with nodal values for the displacement and temperature. The fine scale is a dynamic atomistic unit cell which exists at the quadrature points of the finite elements. The computational advantages of this approach are demonstrated for a few simple examples and it is clearly promising.

4.5.3. Quasi-Static Models for Metals

The contribution of electrons to heat transport is not incorporated into classical MD simulations. As discussed in Section 4.2.6, this contribution is overwhelmingly important in metals, where the thermal conductivity is typically 2–3 orders of magnitude higher than that due to the classical (ballistic) phonon heat transport processes that dominate in insulators (see Table 4-2). One solution to this problem would be to utilize a full quantum mechanics model for the atomistic simulation, except this is computationally too intensive. The alternative solution is to prescribe the thermal properties of the continuum (from experiment) and impose these on the atomistic region. The disadvantage of this is that heat is then primarily transported around the system by locally adding and removing heat by the constant action of strong thermostats. The thermal response of the system is therefore purely homogeneous with no dependence on the system size due to scattering from interfaces and atomistic defects (the electron mean free path is huge compared to the phonon mean free path so size effects could be important at even higher length scales in metals). The advantage in strongly regulating the temperature at all points in the simulation, however, is that the sensitive issues relating to the artificial Kapitza effect and heat transport across the atomistic-continuum interface is no longer a primary issue.

The most widely adopted of these schemes is the Two Temperature Method (TTM). This has mainly been used to consider the effects of laser annealing on voids due to rapid heating (see [7, 113] and references therein) where large atomic rearrangement occurs under the action of high temperatures and temperature gradients. As shown in Figure 4-18, the simulation incorporates two simultaneous models: atomistic and electronic. The atomistic model is an MD simulation. The kinetic energy of the atoms is known as the lattice temperature, T_l. The temperature of the electrons, T_e, evolves in the electronic (continuum) region. The two models interact by the electrons thermalizing the atoms according to

$$\begin{aligned} C_e\dot{T}_e &= \nabla(k_e\nabla T_e) - \lambda(T_e - T_l) + S \\ C_l\dot{T}_l &= \lambda(T_e - T_l) \end{aligned} \tag{4-53}$$

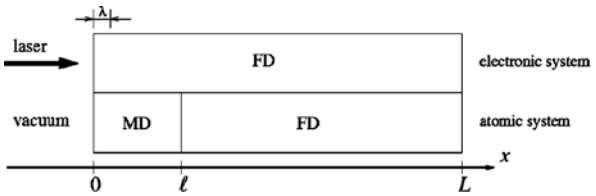


Figure 4-18. Two temperature method (TTM) for heat conduction in metals shows two overlapping regions: atomistic and electronic. The electronic system is simulated using a finite-difference scheme (FD). A laser deposits energy in the electron system according to the source term S in (4-53). The atomic system is treated by a molecular-dynamic scheme down to a depth l and beyond that by a continuum approach (FD). The electronic and atomic systems are coupled by the electron-phonon coupling λ-term in (4-53) (Reprinted (Fig 1, pg 115404-1) with permission from [Schäfer C, Urbassek HM, Zhigilei LV, Phys Rev B, 66]. Copyright by the American Physical Society. http://link.aps.org/abstract/PRB/v66/p115404)

where the standard symbols have been used with the subscripts e and l referring to the properties associated with electrons and atoms respectively, S is a (radiation) heat source (due to the laser) and λ is the strength of the electron-phonon coupling. This final term couples the two simulations and allows the electrons to thermalize the atomistic system (and vice versa to a lesser extent). The lattice temperature is enforced via a velocity rescaling thermostat [7]. A less physical model is proposed by Schall et al. [47] which simply imposes the continuum temperature on the atomistic simulation. This is equivalent to $\lambda \to \infty$ in (4-53) such that $T_l = T_e$ with no characteristic interaction time (apart from the responsiveness of the thermostat). This is used to examine the effects of frictional heating during sliding. The model is also extended to Joule heating [114], whereby the heat source term in (4-53) is determined from the solution for the local electric current determined from an imposed continuum representation.

4.5.4. Dynamic Coarse-Grained Models for Metals

Finally, a multiscale method developed by Liu and Li [108] incorporates many of the features of the CG methodologies discussed previously. As in the bridging scale method [14, 100, 101], the displacement field is decomposed into a fine-scale part and a coarse-scale part. The continuum region extends over the entire region and overlaps the smaller atomistic region. Atoms and coarse-scale nodes are both dynamic and evolve under the action of fine and coarse scale forces. The coarse-scale forces are derived from the quantum CG potential due the Einstein model (4-47). Temperature exists at the coarse-scale and represents the thermal vibrations of the missing atoms where no fine-scale exists. This evolves by the standard continuum relation (4-31). The thermal conductivity is specified using (4-17) where the velocity and mean free path can refer to either phonons or electrons. Material properties such as heat capacity (4-25) and the thermal expansion force (4-48) are derived from the quantum Einstein model. This introduces a deformation-dependence on the heat capacity and hence (4-17) introduces a strain-dependence into the thermal

conductivity. The momenta of the fine-scale is thermalised by the coarse-scale temperature using Nosé-Hoover chains (NHC). Neither scale has the true dynamics of an unthermostatted region, but this is unavoidable for the simulation of heat conduction in metals. Phonons are absorbed at the boundaries of the fine-scale region using a Perfectly Matched Layer [107], as discussed above.

4.5.5. Conclusions

A range of concurrent atomistic-continuum models for finite temperature multiscale simulation in the literature have been discussed. Although zero temperature multiscale methods, such as the quasi-continuum method [9, 11], are becoming more widely used and common tools for the materials modeler, the extension and application of these techniques to finite temperature is still in its infancy. As discussed in Section 4.4.1, the dynamics of CG nodes at finite temperature is now reasonably well established, at least in the prediction of thermophysical properties. The dynamics of a CG node represents the kinetic energy of a single (representative) atom. Given the correct CG mass, the dynamics can also roughly represent the propagation of (low frequency) waves through the CG medium (neglecting thermalisation of the waves due to interaction with higher frequency modes). However, the CG node does not have the dynamics of an individual atom and it does not represent the kinetic energy of the missing (slave) atoms it claims to represent. Only the average position of the CG node is actually a true representation of the representative atom itself. In this respect finite temperature CG dynamics will always be a somewhat abstract feature of multiscale simulation. This raises questions as to how these CG degrees-of-freedom should be treated in a consistent multiscale framework. In isothermal simulation these issues are less important. However, multiscale simulations are almost never exactly isothermal as they are typically subject to internal or external stimuli. As seen in Section 4.5, the elemental temperature must be recorded to represent (at least) the thermal energy of the missing (slave) atoms for inhomogeneous thermal problems in order to conserve thermal energy and model heat transport. In general, the requirements of the CG medium need to be assessed for a particular problem. Does the CG region need to retain dynamic information (about low frequency vibrations) or is a (deformable) quasi-static CG medium, such as those in Sections 4.5.1 and 4.5.3, sufficient? Is the objective to determine the thermal transport properties of a particular device from the simulation? If this is the case, only the phonon-phonon (ballistic) contribution can be determined from classical atomistic models, such as those outlined in Sections 4.5.1 and 4.5.2. This is strictly only applicable to poor electrical conductors (typically carbon and silicon) although, as discussed in Section 4.2.6, approximate methods for estimating the thermal conductivity for metals from such simulations do exist. Imposing the experimental thermal properties of a material on the multiscale model, as in Sections 4.5.3 and 4.5.4, is an attractive option. This is primarily beneficial in the simulation of metals, but it is still applicable to other materials. Taking such strong thermostatting control eliminates many of the issues associated with atomistic/continuum interfaces and

consistency between descriptions. However, such methods are only strictly applicable to the simulation of fairly dramatic situations which are insensitive to changes in the dynamics due to the thermostats. The thermal model is necessarily a bulk model which does not represent the finite size effects due to interfaces and defects seen in Section 4.3. There are clearly a number of different approaches for specific problems and many of these have been addressed, at least in part, by the models reviewed in this chapter. Most of the proposed multiscale modelling approaches discussed here are in their early stages and require more detailed and demanding simulations to fully illustrate the complete competency of the model. No one approach currently addresses all the issues. This is primarily due to the inability of classical MD simulation to fully represent low temperature phonon heat transport and heat transport due to electrons.

ACKNOWLEDGEMENT

The author is grateful for the award of a Royal Academy of Engineering-Leverhulme Trust Senior Research Fellowship 2007–2008.

REFERENCES

1. Allen MP, Tildesley DJ (1987) Computer simulation of liquids. Oxford University Press, New York.
2. Schelling PK, Phillpot SR, Keblinski P (2002) Comparison of atomic-level simulation methods for computing thermal conductivity. Phys Rev B 65:144306.
3. Holian BL, Ravelo R (1995) Fracture simulations using large-scale molecular-dynamics. Phys Rev B 51:11275.
4. Abraham FF, Broughton JQ, Bernstein N, Kaxiras E (1998) Spanning the continuum to quantum length scales in a dynamic simulation of brittle fracture. Europhys Lett 44:783.
5. Lee Y, Park JY, Kim SY, Jun S, Im S (2005) Atomistic simulation of incipient plasticity under Al(111) nanoindetation. Mech Mat 37:1035.
6. Ju Y, Sinnott SB (2004) Constant temperature MD simulations of energetic particle-solid collisions: comparison of temperature control methods. J Comp Phys 200:251.
7. Huang PH, Lai HY (2008) Nucleation and propagation of dislocations during nanopore lattice mending by laser annealing: modified continuum-atomistic modelling. Phys Rev B 77:125408.
8. Curtin WA, Miller RE (2003) Atomistic/continuum coupling in computational materials science. Model Simul Mater Sci Eng 11:R33.
9. Tadmor EB, Ortiz M, Phillips R (1996) Quasicontinuum analysis of defects in solids. Philos Mag A 73:1529.
10. Rudd RE, Broughton JQ (2000) Concurrent coupling of length scales in solid state systems. Phys Stat Sol B 217:5893.
11. Knap J, Ortiz M (2003) Effect of indenter-radius size on Au(001) nanoindentation. Phys Rev Lett 90:226102.
12. Rudd RE, Broughton JQ (2005) CG MD – nonlinear finite elements and finite temperature. Phys Rev B 72:144104.
13. Binder K, Horbach J, Kob W, Paul W, Varnik F (2004) MD simulation. J Phys Condens Matter 16:S429.

14. Liu WK, Karpov EG, Zhang S, Park HS (2004) An introduction to computational nanomechanics and materials. Comput Methods Appl Mech Eng 193:1529.
15. Yip S (2005) Handbook in materials modelling. Springer, Dordrecht .
16. Verlet L (1967) Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules.. Phys Rev 159:98.
17. Janezic D, Orel B (1993) Implicit Runge-Kutta method for MD integration. J Chem Info Comp Sci 33:252.
18. Heino P (2005) Thermal conductivity and temperature in solid argon by nonequilibrium MD simulations. Phys Rev B 71:144302.
19. Baranyai A (2000) Temperature of nonequilibrium steady-state systems. Phys Rev E 62:5989.
20. Braga C, Travis KP (2005) A configurational temperature Nosé-Hoover thermostat. J Chem Phys 123:134101.
21. Andersen HC (1980) MD simulations at constant pressure and/or temperature. J Chem Phys 72:2384.
22. Huang ZX, Tang ZA (2006) Evaluation of momentum conservation influence in non-equilibrium MD methods to compute thermal conductivity. Physica B 373:291.
23. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126:014101.
24. Hoover WG, Ladd AJC, Moran B (1982) High-strain-rate plastic flow via non-equilibrium MD. Phys Rev Lett 48:1818.
25. Evans DJ (1983) Computer "experiment" for nonlinear thermodynamics of Couette flow. J Chem Phys 78:3297.
26. Nosé S (1984) A MD method for simulations in the canonical ensemble. Mol Phys 53:255.
27. Hoover WG (1985) Canonical dynamics: equilibrium phase-space distributions. Phys Rev A 31:1695.
28. Holian BL, Voter AF, Ravelo R (1995) Thermostatted MD: how to avoid the Toda demon hidden in Nosé-Hoover dynamics. Phys Rev E 52:2338.
29. Tuckerman ME, Berne BJ, Martyna GJ (1992). Reversible multiple time scale molecular dynamics. J Chem Phys 97(3):1990–2001.
30. Adelman SA, Doll JD (1976) Generalized Langevin equation approach for atom/solid-surface scattering: general formulation for classical scattering off harmonic solids. J Chem Phys 64:2375.
31. Berendsen HJC (1984) MD with coupling to an external bath. J Chem Phys 81:3684.
32. Leimkuhler BJ, Sweet CR (2005) A Hamiltonian formulation for recursive multiple thermostats in a common timescale. SIAM J Appl Dyn Syst 4:178.
33. Travis KP, Braga C (2008) Configurational temperature control for atomic and molecular systems. J Chem Phys 128:014111.
34. Leimkuhler B, Noorizadeh E, Theil F. (2009) A gentle stochastic thermostat for MD. Journal of Statistical Physics. 135:261.
35. Li X, Weinan W (2007) Variational boundary conditions for MD simulations of crystalline solids at finite temperature: treatment of the thermal bath. Phys Rev B 76:104107.
36. Leimkuhler B, Legoll F, Noorizadeh E (2008) A temperature control technique for nonequilibrium molecular simulation. J Chem Phys 128:074105.
37. Lepri S, Livi R, Politi A (2003) Thermal conduction in classical low-dimensional lattices. Phys Rep 377:1.
38. Heino P (2007) Thermal conduction simulations in the nanoscale. J Comput Theor Nanosci 4:896.
39. Murthy JY, Narumanchi SVJ, Pascual-Gutierrez JA, Wang T, Ni C, Mathur SR (2005) Review of multiscale simulation in submicron heat transfer. Int J Multiscale Comput Eng 3:5.
40. Kittel C (2005) Introduction to solid state physics. John Wiley and Sons, Inc, New York, 8th edition.

41. Griffiths, DJ (2005) Introduction to quantum mechanics. Pearson Education Inc., Upper Saddle River, NJ, 2nd ed.
42. Jolley K, Gill SPA (2009) Modelling transient heat conduction in solids at multiple length and time scales: a coupled non-equilibrium MD/continuum approach. J Comp Phys 228:7412.
43. Sinha S, Goodson KE (2005) Review: multiscale thermal modelling in nanoelectronics. Int J Multiscale Comput Eng 3:107.
44. Wang JS, Wang J, Lü JT (2008) Quantum thermal transport in nanostructures. Eur Phys J B 62:381.
45. Wang J (2007) Quantum thermal transport from classical MD. Phys Rev Lett 99:160601.
46. Sinha S, Goodson KE (2006) Thermal conduction in sub-100 nm transistors. Microelectron J 37:1148.
47. Schall JD, Padgett CW, Brenner DW (2005) Ad hoc continuum-atomistic thermostat for modelling heat flow in MD simulation. Mol Simulat 31:283.
48. Chantrenne P, Raynaud M, Baillis D, Barrat JL (2003) Study of phonon heat transfer in metallic solids from MD simulations. Microscale Thermophys Eng 7:117.
49. Yuan SP, Jiang PX (2006) Thermal conductivity of small nickel particles. Int J Thermophys 27:581.
50. Webb III EB, Zimmerman JA, Seel SC (2008) Reconsideration of continuum thermomechanical quantities in atomic scale simulations. Math Mech Solids 13:221.
51. Desgranges C, Delhommelle J (2008) Molecular simulation of transport in nanopores: application of the transient-time correlation function formalism. Phys Rev E 77:027701.
52. Jund P, Jullien R (1999) MD calculation of the thermal conductivity of vitreous silica. Phys Rev B 59:13707.
53. Stevens RJ, Zhigilei LV, Norris PM (2007) Effects of temperature and disorder on thermal boundary conductance at solid-solid interfaces: nonequilibrium MD simulations. Int J Heat Mass Transfer 50:3977.
54. Park HS, Karpov E, Liu WK (2004) A temperature equation for coupled atomsitic/continuum simulations. Comput Methods Appl Mech Eng 193:1713.
55. Prasher R, Tong T, Majumdar A (2007) Diffraction-limited phonon thermal conductance of nanoconstrictions. Appl Phys Lett 91:143119.
56. Bhowmick S, Shenoy VB (2006) Effect of strain on the thermal conductivity of solids. J Chem Phys 125:164513.
57. Terao T, Müller-Plathe F (2005) A nonequilibrium MD method for thermal conductivities based on thermal noise. J Chem Phys 112:081103.
58. Daly BC, Maris HJ, Imamura K, Tamura S (2002) MD calculation of the thermal conductivity of superlattices. Phys Rev B 66:024301.
59. Hulse RJ, Rowley RL, Wilding WV (2005) Transient nonequilibrium molecular dynamic simulation of thermal conductivity: 1. Simple fluids. Int J Thermophys 26:1.
60. Chen G, Borca-Tasciuc D, Yang RG (2004) Nanoscale heat transfer. Encyclopedia of nanoscience and nanotechnology. American Scientific publishers, Valencia, CA.
61. Cahill DG, Ford WK, Goodson KE, Mahan GD, Majumdar A, Maris HJ, Merlin R, Phillpot SR (2003) Nanoscale thermal transport. J Appl Phys 93:793.
62. Srinivasan S, Miller RS (2007) On parallel nonequilibrium MD simulations of heat conduction in heterogeneous materials with three-body potentials: Si/Ge superlattice. Numer Heat Transfer B 52:297.
63. Yang YW, Liu XJ, Yang JP (2008) Nonequilibrium MD simulation for size effects on thermal conductivity of Si nanostructures. Mol Simulat 34:51.
64. Tang Q (2004) A MD simulation: the effect of finite size on the thermal conductivity in a single crystal silicon. Mol Phys 102:1959.

65. Lee JH, Grossman JC, Reed J, Galli G (2007) Lattice thermal conductivity of nanoporous Si: MD study. Appl Phys Lett 91:223110.
66. Heino P (2007) Dispersion and thermal resitivity in silicon nanofilms by MD. Eur Phys J B 60:171.
67. Ponomareva I, Srivastava D, Menon M (2007) Thermal conductivity in thin silicon nanowires: phonon confinement effect. Nano Lett 7:1155.
68. Balandin A, Wang KL (1998) Significant decrease of the lattice thermal conductivity due to phonon confinement in a free-standing semiconductor quantum well. Phys Rev B 58:1544.
69. Segal D, Nitzan A (2003) Thermal conductance through molecular wires. J Chem Phys 119:136840.
70. Zhong H, Lukes JR (2006) Interfacial thermal resistance between carbon nanotubes: MD simulations and analytical thermal modelling. Phys Rev B 74:125403.
71. Gu Y, Chen Y (2007) Thermal conductivities of single-walled carbon nanotubes calculated from the complete phonon dispersion relations. Phys Rev B 76:134110.
72. Che J, Çağin T, Goddard III WA (2000) Thermal conductivity of carbon nanotubes. Nanotechnology 11:65.
73. Cao BY, Hou QW (2008) Thermal conductivity of carbon nanotubes embedded in solids. Chin Phys Lett 25:1392.
74. Bi K, Chen Y, Yang J, Wang Y, Chen M (2006) MD simulation of thermal conductivity of single-walled carbon nanotubes. Phys Lett A 350:150.
75. Pan RQ, Xu ZJ, Zhu ZY (2007) Length dependence of thermal conductivity of single-walled carbon nanotubes. Chin Phys Lett 24:1321.
76. Zhang W, Zhu Z, Wang F, Wang T, Sun L, Wang Z (2004) Chirality dependence of the thermal conductivity of carbon nanotubes. Nanotechnology 15:936.
77. Tang Q, Yao Y (2006) The Kapitza resistance across grain boundary by MD simulation. Nanoscale Microscale Thermophys Eng 10:387.
78. Watanabe T, Ni B, Phillpot SR, Schelling PK, Keblinski P (2007) Thermal conductance across grain boundaries in diamond from MD simulation. J Appl Phys 102:063503.
79. Alvarez-Quintana J, Alvarez X, Rodriguez-Viejo J, Jou D, Lacharmoise PD, Bernardi A, Goñi AR, Alonso MI (2008) Cross-plane thermal conductivity reduction of vertically uncorrelated Ge/Si quantum dot superlattices. Appl Phys Lett 93:013112.
80. Mingo N, Yang L, Li D, Majumdar A (2003) Predicting the thermal conductivity of Si and Ge nanowires. Nano Lett 3:1713.
81. Li D, Wu Y, Kim P, Shi L, Yang P, Majumdar A (2003) Thermal conductivity of individual silicon nanowires. Appl Phys Lett 83:2934.
82. Hone J (2004) Carbon nanotubes: thermal properties. Dekker encyclopaedia of nanoscience and nanotechnology 603. Marcel Dekker Inc, New York.
83. Chantrenne P, Barrat JL (2004) Analytical model for the thermal conductivity of nanostructures. Superlattice Microstruct 35:173.
84. Biercuk MJ, Llaguno MC, Radosavljevic M, Hyun JK, Johnson AT (2002) Carbon nanotube composites for thermal management. Appl Phys Lett 80:2767.
85. Lukes JR, Zhong H (2007) Thermal conductivity of individual single-walled carbon nanotubes. J Heat Transfer 129:705.
86. Keblinski P, Schelling PK (2005) Comment on "Thermal contraction of carbon fullerenes and nanotubes". Phys Rev Lett 94:209702.
87. Kwon YK, Berber S, Tomanek D (2097) Reply to comment on "Thermal contraction of carbon fullerenes and nanotubes". Phys Rev Lett 94:02.
88. Xiao SP, Belytschko T (2004) A bridging domain method for coupling continua with MD. Comput Methods Appl Mech Eng 193:1645.

89. Liu J, Chen S, Nie X, Robbins MO (2007) A continuum-atomistic simulation of heat transfer in micro- and nano- flows. J Comp Phys 227:279.
90. Flekkøy EG, Delgado-Buscalioni R, Coveney PV (2005) Flux boundary conditions in particle simulations. Phys Rev E 72:026703.
91. Werder T, Walther JH, Koumoutsakos P (2005) Hybrid atomistic-continuum method for the simulation of dense fluid flows. J Comp Phys 205:373.
92. Lu G, Tadmor EB, Kaxiras E (2006) From electrons to finite elements: a concurrent multiscale approach for metals. Phys Rev B 73:024108.
93. LeSar R, Najafabadi R, Srolovitz DJ (1989) Finite-temperature defect properties from free-energy minimization. Phys Rev Lett 63:624.
94. Wu ZB, Diestler DJ, Feng R, Zeng XC (2003) Coarse-graining description of solid systems at nonzero temperature. J Chem Phys 119:8013.
95. Gill SPA, Jia Z, Leimkuhler B, Cocks ACF (2006) Rapid thermal equilibration in CG MD. Phys Rev B 73:184304.
96. Dupuy LM, Tadmor EB, Miller RE, Phillips R (2005) Finite-temperature quasicontinuum: MD without all the atoms. Phys Rev Lett 95:060202.
97. Curtarolo C, Ceder G (2002) Dynamics of an inhomogeneously CG multiscale system. Phys Rev Lett 88:255504.
98. Kirkwood JG (1935) Statistical mechanics of fluid mixtures. J Chem Phys 3:300.
99. Cai W, de Koning M, Bulatov VV, Yip S (2000) Minimizing boundary reflections in coupled-domain simulations. Phys Rev Lett 85:3213.
100. Liu WK, Park HS, Qian D, Karpov EG, Kadowaki H, Wagner GJ (2006) Bridging scale methods for nanomechanics and materials. Comput Methods Appl Mech Eng 195:1407.
101. Park HS, Liu WK (2004) An introduction and tutorial on multiple-scale analysis in solids. Comput Methods Appl Mech Eng 193:1733.
102. Weinan E, Huang Z (2002) A dynamic atomistic-continuum method for the simulation of crystalline materials. J Comp Phys 182:234.
103. Karpov E, Park HS, Liu WK (2007) A phonon heat bath approach for the atomistic and multiscale simulation of solids. Int J Numer Meth Eng 20:351.
104. Qu S, Shastry V, Curtin WA, Miller RE (2005) A finite temperature dynamic coupled atomisitic/discrete dislocation method. Model Sim Mater Sci Eng 13:1101.
105. Shilkrot LE, Miller RE, Curtin WA (2004) Multiscale plasticity modelling: coupled atomistic and discrete dislocation mechanics. J Mech Phys Solids 52:755.
106. Tang S (2008) A finite difference approach with velocity interfacial conditions for multiscale computations of crystalline solids. J Comp Phys 227:4038.
107. Li S, Liu X, Agrawal A, To AC (2006) Perfectly matched multiscale simulations for discrete lattice systems: extension to multiple dimensions. Phys Rev B 74:045418.
108. Liu X, Li S (2007) Nonequilibrium multiscale computational model. J Chem Phs 126:124105.
109. Namilae S, Nicholson DM, Nukala PKVV, Gao CY Ostesky YN, Keffer DJ (2007) Absorbing boundary conditions for MD and multiscale simulation. Phys Rev B 76:144111.
110. Weinan E, Enquist B, Li X, Ren W, Vanden-Eijnden E (2007) Heterogeneous multiscale methods: a review. Commun Comput Phys 2:367.
111. Li X, Weinan, W (2005) Multiscale modelling of the dynamics of solids at finite temperature. J Mech Phys Solids 53:1650.
112. Fish J, Chen W, Li R (2007) Generalized mathematical homogenization of atomistic media at finite temperatures in three dimensions. Comput Methods Appl Mech Eng 196:908.
113. Schäfer C, Urbassek HM, Zhigilei LV (2002) Metal ablation by picosecond laser pulses: a hybrid simulation. Phys Rev B 66:115404.
114. Padgett CW, Brenner DW (2005) A continuum-atomistic method for incorporating Joule heating into classical MD simulations. Mol Simulat 31:749.

CHAPTER 5

A MULTISCALE METHODOLOGY TO APPROACH NANOSCALE THERMAL TRANSPORT

ISHWAR K. PURI[1] AND SOHAIL MURAD[2]

[1]*Department of Engineering Science and Mechanics, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24061, USA,e-mail: ikpuri@vt.edu*

[2]*Department of Chemical Engineering, University of Illinois at Chicago, Chicago, IL, 60607, USA, e-mail: murad@uic.edu*

Abstract: The contact resistance problem between dissimilar or bonded substrates is particularly important at the nanoscale, since the length scales associated with the structures and energy carriers become comparable. We provide a basic understanding of nanoscale thermal properties, focusing on nanoscale composition and surface structure effects on local and bulk thermal properties, and discuss how surface modifications can create novel materials and structures that have tunable thermal properties. Since nanoscale flows are typically part of larger scale systems and we are confronted with an inherently multiscale problem, a multiscale approach is required to integrate atomistic simulations with computational methods suitable for flow phenomena at larger scales. We begin by describing how nanoscale thermal transport can be investigated using molecular dynamics (MD) simulations for ideal (defect-free) materials, with defects, and with simpler (solid-solid, solid-liquid, solid-vapor, etc.) and more complex (solid-liquid-solid, solid-liquid-vapor, liquid-vapor-liquid) material contacts. Next, we describe how the mesoscale lattice Boltzmann method (LBM) can be used to model thermal transport. Then, we describe a hybrid model that couples MD with LBM. Finally, we provide examples of several problems suitable for the multiscale modeling of thermal transport

Keywords: Multiscale methods, Molecular dynamics, Lattice Boltzmann method, Nanoscale transport, Heat transfer

5.1. INTRODUCTION

The contact resistance problem [1] between dissimilar or bonded substrates is particularly important at the nanoscale, since the length scales associated with the structures and energy carriers become comparable. The thermal management of electronic packaging of thermo-electrics [2], nanofabrication [3], and heat transfer though microchannels [4] has prompted atomic-level investigations of this resistance.

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 135–150.
DOI 10.1007/978-1-4020-9785-0_5,

5.1.1. Interfacial Resistance

When thermal energy is transported through a solid-fluid interface of area A, it produces a temperature discontinuity across the boundary [5]. If the heat flow Q' across the interface is small, the temperature difference across the interface ΔT is thought to be proportional to it. The effective thermal resistance $\Delta T/Q'$ is typically expressed as the Kapitza resistance [6–10], $R = A\ \Delta T/\ Q'$ (m^2 K/W). The first measurements of R were made by Kapitza for metal surfaces suspended in He II (the superfluid phase of ^{4}He) in the temperature range between 1.6 and 2.12 K. These and other similar experiments involving other substances and temperatures suggest that $R \propto T^{\alpha}$. The Kapitza resistance can be represented as the inverse interfacial conductance [11], i.e., $R = 1/G = AdT/Q'$, or in terms of the Kapitza length $l = R\lambda$. Here, λ denotes the thermal conductivity of the bulk medium and l is the equivalent width of the bulk medium over which the temperature drop is the same as over the interface [12, 13]. For a normal liquid-solid interface, l is on the order of few molecular diameters [12].

Investigations of this conductance at the nanoscale have been generally restricted to inert defect-free ideal interfaces that do not consider dissimilar molecules placed on the interface, interfacial defects, or the application of external fields. The role of defects, e.g., when carbon nanotubes (CNTs) bond to other materials, have catalyst nanoparticles at their tips, or contain atomic vacancies and array misalignments, is also important. (For instance, the measured thermal conductivity along the axial direction of multiwall CNTs can be much smaller than theoretical predictions for idealized single-walled CNTs [14]).

5.1.2. Phonon Behavior Through Acoustic Waves

Phonons are quanta of lattice vibrational energy. They play a major role in determining the thermal and electrical resistances of a material. The effects of the interfacial thermal (Kapitza) can be explained in the light of phonon scattering [11, 15]. The thermal conductivity of an interface depends on electron-phonon and phonon-phonon interactions as well as electron and phonon scatterings [16]. Calculating phonon interactions that limit thermal conductivity is an undoubtedly complex problem but can be mechanically examined by considering phonon-phonon interactions that lead to acoustic waves [17].

Like other transport and thermodynamic properties, the thermal behavior of nanostructured interfaces, e.g., in nanoelectronic devices or across nanometer scale point like constrictions, cannot be simply inferred by extrapolating bulk behavior to the smaller scales. Nanoscale thermal transport differs from transport in bulk materials because the mean free path for phonon scattering can be large compared to device dimensions and lead to interesting physics such as quantum thermal phenomena [18]. When the thermal transport dimension is comparable to the dominant phonon wavelength, the interface thermal resistance, or Kapitza resistance, is considerably lower than the calculated ballistic resistance because bulk phonon dispersion and bulk potential are no longer accurate [19].

The thermal resistance can be explained through the acoustic mismatch model, which assumes it to arise from the large impedance to the passage of thermal phonons across a solid-fluid interface. The acoustic impedance governing the transmission and reflection of these phonons is the product ρc of the density and sound velocity. It can be many orders of magnitude greater for a solid than a fluid. Due to an acoustic mismatch, a large fraction of the phonons impinging upon such an interface from both sides are unable to pass through it. This model typically overpredicts the thermal resistance and is considered as an upper bound on it. The diffuse mismatch model provides a lower bound on R. It assumes nonspecular behavior at the interface, i.e., that all phonons are diffusively scattered. The phonon transmission probability is again related to a mismatch, in this case between the different densities of the solid-fluid states [7].

5.1.3. Strategies to Modulate the Interfacial Resistance

The nature of a nanoscale interface has an important influence on thermal transport [6, 9, 10]. An understanding of phonon behavior suggests several strategies to reduce the mismatch at the interface, thus decreasing R and increasing Q'/A, as follows. (1) Since the product ρc increases with pressure, one approach could be to simply increase the fluid pressure to facilitate better acoustic matching, thus lowering R. (2) The interface could be made more solid-like by adsorbing and ordering additional fluid molecule layers [6, 7, 10, 20], e.g., by making the surface more hydrophilic [21, 22]. For both cases, the impedance of a dense solid-like layer on the fluid side would be intermediate between the corresponding values for the solid and fluid. Once formed, an open question is if these properties are relatively pressure independent [23].

5.1.4. Role of Surface Modifications

Surface structure has a profound effect on a material's wettability and, thus on its thermal transport characteristics. We address this with a brief and rudimentary discussion of intermolecular and surface forces [24]. Intermolecular forces are responsible for the cohesion of molecules in solids and liquids. Water is different from many liquids in that it is a polar molecule that has strong hydrogen bonds (∼10–15 kJ/mol) between the water molecules. Other non-polar liquids are held together by weaker van der Waals forces (∼1 kJ/mol), while hard solids are held together by exceptionally strong covalent or metallic bonds (∼500 kJ/mol). Nonpolar substances are generally incapable of strongly interacting (associating) with water. These materials are hydro*phobic,* as water does not "wet" their surfaces. A familiar example is wax upon which water beads up, or certain forms of carbon nanostructures deposited upon Si [25]. On the other hand hydro*philic* materials are either polar or contain electronegative atoms capable of interacting strongly with water.

As another example, carbon films can be deposited [26] in the form of nanotubes [27, 28] and nanobeads, e.g., on silicon wafers, in order to modify the substrate

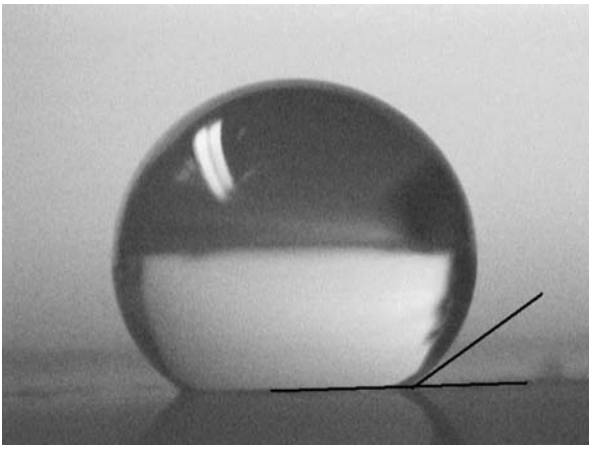


Figure 5-1. Images of ≈2 mm water droplets placed on a Si disc containing a carbon nanobead deposit [25]

surface wettability, as shown in Figure 5-1 [25]. Carbon nanostructures, particularly nanotubes, also provide a test bed for synergistic simulations and measurements [29] such as of their thermal transport properties [30–32], flow within and outside the structures [33–36], energy exchange with fluids [37, 38], interfacial thermal transport between adjacent nanostructures [39], and effects of the boundary resistance [40] and chemical bonding (e.g., between the matrix and fiber on thermal transport in case of nanotube composites [41]). Some of this research has been motivated by the potential use of carbon nanotubes in composite materials to enhance their thermal conductivity [42–45]. Thus, we note that there are many interesting configurations that lend themselves to simulations to elucidate the nanoscale interfacial thermal properties between dissimilar materials.

5.2. CONTINUUM LIMITS

In the absence of interfaces, nanoscale transport can follow the regular continuum-scale physical relations, even for nanoscale segments containing only 10 molecules in each spatial direction [21, 46–48]. However, the nature of an interface and the material adjacent to it have an important influence on the associated nanoscale transport phenomena [9, 10, 21, 36, 46, 48]. For instance, transport in nonpolar fluids deviates significantly from the predictions of continuum theories because of thermal fluctuations [49–51]. In contrast, heat and mass transport in strongly polar fluids that have significant intermolecular interactions (such as the high dipole moment and hydrogen bonding in water) can be much different. Here, molecules lose memory of their initial or previous velocities, i.e., their local energies, much more rapidly after collisions occur, thereby significantly attenuating the role of thermal fluctuations. Therefore, the influence of polarity and the action of external fields (e.g., electric or

magnetic) across interfaces on nanoscale transport phenomena must be investigated more closely [46, 48].

5.3. MULTISCALE INVESTIGATIONS

5.3.1. Atomistic and Multiscale Simulations

Nanoscale thermal transport can be investigated using molecular dynamics (MD) simulations for ideal materials, those with defects, and with simpler (solid-solid, solid-liquid, solid-vapor, etc.) and more complex (solid-liquid-solid, solid-liquid-vapor, liquid-vapor-liquid) material contacts. Examples of such simulations include examinations of liquid thermal conductivity [52], thermal transport in the so-called "nanofluids" [53] (a perplexing misnomer), or Si [54], and the effects of grain size [55, 56] and phonon scattering [57] on thermal resistance across grain boundaries. Nonequilibrium (NEMD) simulations that rely on the imposition of temperature gradients within the system are used to investigate thermal transport across solid-solid [58] and liquid-liquid interfaces [59], across thin films [60–63], and to determine transport properties such as the interfacial (Kapitza) thermal resistance [12]. Their success depends on how well atomic interactions are modeled and how long the simulation lasts.

MD simulation methodology can handle both equilibrium and nonequilibrium, and unsteady and steady state problems. It can account for the implicit effects of acoustic waves. Temperature is maintained in such simulations by a Gaussian thermostat where molecules have an initial Gaussian velocity distribution corresponding to the local temperature. The rate of heat transfer can be obtained from the energy supplied by the heat source (higher temperature) that is removed by the heat sink (lower temperature). One method that we have used uses a fifth-order Gear predictor–corrector algorithm for translational motion; a fourth-order predictor–corrector algorithm for rotational motion is used by employing the quaternion method [36, 46, 48]. We can investigate unusual effects, such of electric fields [36, 48], on the thermal transport rate for both atomistic simulations and the multiscale hybrid simulations. We are also able to simulate complex interactions such as the impingement of nanojets on inert and variously hydrophobic surfaces (e.g., shown in Figure 5-2).

Nanoscale flows are typically part of larger scale systems, e.g., when nanofluidic channels interface with microfluidic domains, and we are confronted with an inherently multiscale problem. Hence, a multiscale approach is required to integrate atomistic simulations with computational methods suitable for flow phenomena at larger scales [64]. Multiscale analyses bridge length and time scales [65] and suitably incorporate boundary conditions across them [66]. The boundary conditions for realistic surfaces are often unknown, but become increasingly important as the system size decreases into the micrometer and nanometer scales. These unknown boundary conditions have led to intense interest in the development of hybrid methods that can determine interfacial boundary conditions atomistically while solving continuum equations away from the interface [67].

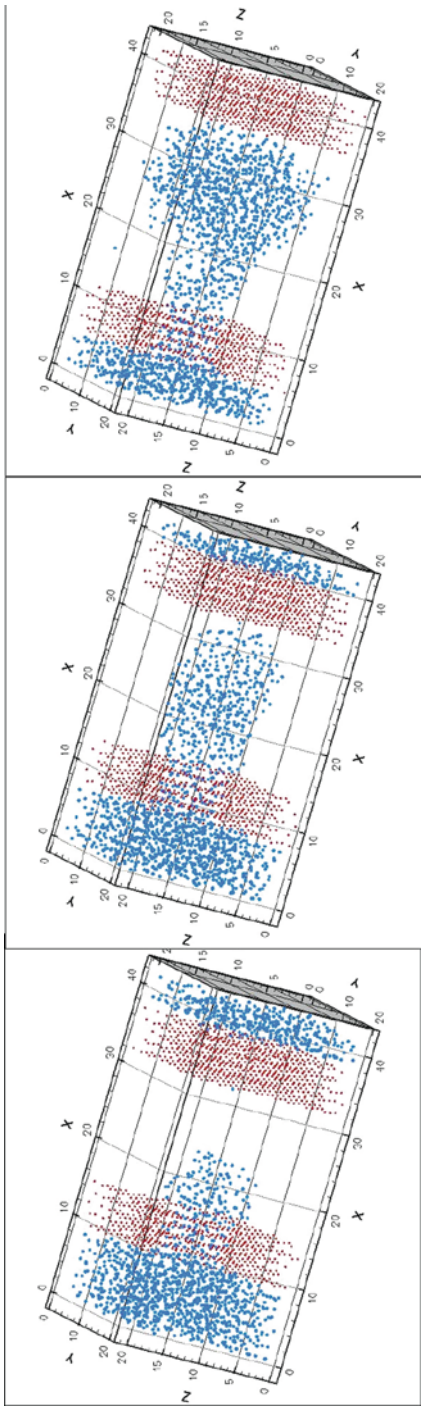


Figure 5-2. Consecutive images from a simulation depicting the evolution of a nanojet and its impingement on a wall. The flow occurs from *left* to *right*. The nanojet emerges from the left orifice while the opposite (*right*) wall serves as a stagnation surface. The nanojet fluid is colored *blue* and the membrane molecules are *red*. Results are presented for various time steps [21]

Despite the success of atomistic simulations, their limitations in accessible length and time scales are stringent and allow only for the analysis of elementary systems for rather short times. To illustrate these limitations, consider that the time step dt in an MD simulation is dictated by the fastest frequency one needs to resolve. For a simulation of pure water, $dt = 2$ fs when models with fixed O–H bonds and H–O–H angles are used; in other words, 500 million time steps are required for 1 μs of simulation time. With the optimistic assumption that the execution of a single time step takes 0.1 s, a total of some 19 months of CPU time is required [68]. Since full atomistic simulations are prohibitively expensive, hybrid atomistic–continuum simulations are necessary to study large systems for reasonable times. Here we describe a hybrid model coupling the molecular dynamics (MD) and lattice Boltzmann method (LBM). Thus, it is possible to take advantage of the mesoscopic modeling inherent in LBM to allow for a broader geometric flexibility than the one allowed for in finite volume solver [64].

5.3.2. Molecular Dynamics (MD) Simulations

The most fundamental theoretical techniques that can provide realistic results are molecular simulation methods such as MD or Monte Carlo simulations. These methods have been widely used for investigating transport phenomenon at the nanoscale [47, 48]. MD is a deterministic method that solves for the equation of motion of individual molecules and looks at the time evolution of the system. The position of each atom/molecule is determined by solving the equations of motion

$$\frac{d}{dt}\mathbf{r}_i = \mathbf{v}_i \quad \text{and} \quad m_i \frac{d}{dt}\mathbf{v}_i = \mathbf{F}_i = -\sum_{j \neq i} \nabla U(r_{ij})$$

where

$$U(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + U_b\left(r_w\right). \tag{5-1}$$

Here, $\mathbf{r}_\mathrm{i}$ denotes the position vector of an atom, $\mathbf{v}_\mathrm{i}$ the corresponding velocity vector, $\mathbf{r}_\mathrm{ij}$ the relative position between sites i and j, and U the Lennard-Jones potential. The term $U_b(r_w)$ is the boundary potential and accounts for the interaction of the boundary region with the surrounding medium. MD is particularly advantageous in the case of solids, higher density fluids, such as gases under high pressure, or liquids, and can be used to calculate dynamic properties. The accuracy of the results predicted by MD depends on the accuracy of the potential function that is used. Therefore, one needs to be careful in selecting the potential function such as the representation for U.

Simulations can be based on the following steps: (1) *Establish the initial configuration or geometry of a nanostructure*. This will vary with the specific geometry of interest such as the molecular details and crystal structure of a solid-liquid interface;

and (2) *Prescribe the intermolecular interactions*. The interactions between a solid crystal and the liquid phase must be modeled as well as the liquid-liquid or solid-solid interactions. In the case of a transition metal, e.g., iron, it is possible to use the embedded atom model (EAM) to accurately model the solid phase. The liquid phase can be modeled with a Lennard-Jones potential.

MD is also a classical approach in which each vibrational mode is equally excited and it does not consider electron contributions to heat conduction. Thus, in the context of phonon transport it is rigorously applicable only to nonmetallic solids above their Debye temperatures T_D [69]. For large enough interface temperatures T when $T_D \neq f(T)$, the thermal resistance is inversely proportional to the interface temperature, i.e., $R \propto T^{-\alpha}$ [70].

5.3.3. Thermal Lattice Boltzmann Method (LBM)

The fundamental concept defining the LBM is the construction of simplified kinetic models that incorporate the essential physics of microscopic and mesoscopic processes so that the averaged macroscopic properties obey the desired macroscopic equations [71–73]. The Boltzmann equation with the Bhatnagar-Gross-Krook (BGK) approximation is [74]

$$\left(\frac{\partial}{\partial t} + \mathbf{e}\cdot\nabla_{\mathbf{r}} + \mathbf{a}\cdot\nabla_{\mathbf{e}}\right) f(\mathbf{r},\mathbf{e},t) = -\frac{f - f_{\mathrm{eq}}}{\tau},$$

where $f(\mathbf{r}, \mathbf{e}, t)$ is a one-particle probability distribution function defined such that $[f(\mathbf{r}, \mathbf{e}, t).d^3\mathbf{r}.d^3\mathbf{e}]$ is the number of particles that at time t are located within a phase-space control element $[d^3\mathbf{r}.d^3\mathbf{e}]$ about $\mathbf{r}$ and $\mathbf{e}$ (where $\mathbf{r}$ is the particle's coordinate in physical space and $\mathbf{e}$ is the particle's discrete velocity) [75]. Here, $\mathbf{a}$ denotes the external force per unit mass acting on the particle. The last term in the BGK approximation of the Boltzmann equation represents the collision between the two particles and is the BGK collision operator [76]. The equilibrium distribution f_{eq} is generally taken to be the Maxwell-Boltzmann distribution for molecules for which $\nabla_{\mathbf{e}} f \approx \nabla_{\mathbf{e}} f_{\mathrm{eq}} = \frac{\mathbf{e}-\mathbf{u}}{RT} f_{\mathrm{eq}}$. Thus,

$$\frac{\partial f}{\partial t} + \mathbf{e}\cdot\nabla_{\mathbf{r}} f = -\frac{f - f_{\mathrm{eq}}}{\tau} + \frac{\mathbf{a}\cdot(\mathbf{e}-\mathbf{u})}{RT} f_{\mathrm{eq}}.$$

This classical view must be modified to model thermal transport. The thermal energy distribution model uses a new distribution function to simulate the temperature field, whereas the density and velocity fields are simulated using a modified density distribution function. The modified density distribution function and new thermal energy distribution function satisfy the following equations (where time and space are discretized along a discretized finite set of velocity directions $\mathbf{e}_\alpha$) [77],

$$\tilde{f}_\alpha(\mathbf{r}_i + \mathbf{e}_\alpha\delta t, t+\delta t) - \tilde{f}_\alpha(\mathbf{r}_i, t) = -\frac{\delta t}{\tau_v + 0.5\delta t}\left[\tilde{f}_\alpha(\mathbf{r}_i, t) - f_\alpha^{\mathrm{eq}}(\mathbf{r}_i, t)\right] + \frac{\tau_v F_\alpha \delta t}{\tau_v + 0.5\delta t}, \tag{5-2a}$$

and

$$\tilde{g}_\alpha(\mathbf{r}_i + \mathbf{e}_\alpha \delta t, t + \delta t) - \tilde{g}_\alpha(\mathbf{r}_i, t) = -\frac{\delta t}{\tau_c + 0.5\delta t}\left[\tilde{g}_\alpha(\mathbf{r}_i, t) - g_\alpha^{\mathrm{eq}}(\mathbf{r}_i, t)\right] + \frac{\tau_c f_\alpha(\mathbf{r}_i, t)\, q_\alpha \delta t}{\tau_c + 0.5\delta t} \tag{5-2b}$$

where, $F_\alpha = \frac{\mathbf{a}\cdot(\mathbf{e}_\alpha - \mathbf{u})\delta t}{RT} f_\alpha^{\mathrm{eq}}(\mathbf{r}_i, t)$, $\mathbf{a}$ the force per unit mass and

$$q_\alpha = (\mathbf{e}_\alpha - \mathbf{u}) \cdot \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{e}_\alpha \cdot \nabla)\,\mathbf{u}\right].$$

Relaxation times are related to the fluid viscosity and thermal diffusivity through the relations $\nu = (1/3)\tau_\nu (\delta x/\delta t)^2$ and $\chi = (1/6)\tau_c (\delta x/\delta t)^2$, respectively. Fluid density, momentum, and energy are calculated from the moments of the distribution function, $\rho = \sum_\alpha \tilde{f}_\alpha$, $\rho\mathbf{u} = \sum_\alpha \tilde{f}_\alpha \mathbf{e}_\alpha + \rho\mathbf{a}\delta t/2$, and $\rho\varepsilon = \sum_\alpha \tilde{g}_\alpha - (\delta t/2)\sum_\alpha f_\alpha q_\alpha$. Pressure can be obtained from the relation $p = (1/3)\rho c^2$.

The original density distribution function f_α and thermal distribution function g_α are used for the boundary conditions. These original unmodified distribution functions are related to the modified ones through the following equations,

$$\tilde{f}_\alpha = f_\alpha + (\delta t/2\tau_\nu)\left(f_\alpha - f_\alpha^{\mathrm{eq}}\right) - (\delta t/2)F_\alpha,$$

and

$$\tilde{g}_\alpha = g_\alpha + (\delta t/2\tau_c)\left(g_\alpha - g_\alpha^{\mathrm{eq}}\right) + (\delta t/2) f_\alpha q_\alpha.$$

The bounce-back rule for the nonequilibrium distribution function [78] can be used for the boundary condition. The density distribution function at the boundary should satisfy the condition $f_\alpha^{\mathrm{neq}} = f_\beta^{\mathrm{neq}}$, where $\mathbf{e}_\alpha$ and $\mathbf{e}_\beta$ are in opposite directions. Thus, the thermal boundary condition is [77]

$$g_\alpha^{\mathrm{neq}} - \mathrm{e}_\alpha^2 f_\alpha^{\mathrm{neq}} = -\left(g_\beta^{\mathrm{neq}} - \mathrm{e}_\beta^2 f_\beta^{\mathrm{neq}}\right).$$

5.3.4. Hybrid Multiscale Methodology

In a hybrid multiscale simulation, one key issue is the appropriate coupling of length and time scales for the two descriptions. Significant progress has been made in solving both problems in the case of rarefied gas flows [79, 80]. However, for dense fluids the situation is more complex since the atomistic description involves interacting particles. Two classes of coupling schemes for dense fluids have been proposed, one based on direct flux exchange [81–83] and the second on the Schwarz alternating method [84, 85]. Direct flux exchange schemes decouple length but not time scales while the Schwartz method finds a consistent solution iteratively in both atomistic and continuum domains.

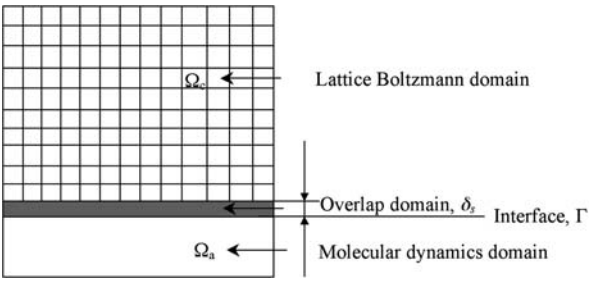


Figure 5-3. The multiscale computation domain

The Schwarz method avoids the direct imposition of fluxes but ensures flux continuity under the assumption that the transport coefficients of the two descriptions match one another in the overlap region. Here, the computational domain is decomposed into two overlapping regions: an atomistic region described by MD and a continuum region described by LBM as shown in Figure 5-3. A Schwartz iteration consists of computing the continuum velocity field $\mathbf{u}_c(t_c)$ and temperature field $T_c(t_c)$ with boundary conditions set by the previous atomistic cycle $\mathbf{u_a}(t_c{-}1)$ and $T_a(t_c{-}1)$ and by an external boundary condition that depends on the system configuration. Then, $\mathbf{u}_c(t_c)$ and $T_\mathbf{c}(t_c)$ are used for setting up the boundary condition for computing $\mathbf{u_a}(t_c)$ and $T_\mathbf{a}(t_c)$. The iteration runs until convergence.

5.3.5. Coupling MD and LBM

The coupling between MD and LBM leads to a number of challenges that must be addressed including (1) the presence of periodic boundary conditions in the MD simulations, (2) the sampling of the MD solution over small regions that will serve as a boundary condition for the continuum, and in turn, (3) the imposition of the continuum boundary conditions onto each cell of the atomistic domain. The presence of periodicity in MD simulations requires a mechanism to compensate for particles that exit the boundary. Simplified models such as elastic boundaries and/or particle reinsertion often result in a strong density gradient and unphysical system behavior.

A boundary force can be imposed to ensure a correct mean pressure for at the MD/LBM interface that will minimize local disturbances in flow quantities like pressure and density. A number of boundary force models have been employed in hybrid schemes. The simulations can use a model [68] that maintains constant density in the atomistic portion of the domain. This boundary model accounts for the local structure of the fluid described by a suitable radial function $g(r)$. The force components normal to the wall and the potential energy contributions weighted by $g(r)$ can be integrated over the part of the cutoff sphere that lies outside the atomistic domain, as shown in Figure 5-4. The integration can, for instance, be performed in polar coordinates, where z is normal to the boundary and x denotes the radial direction as

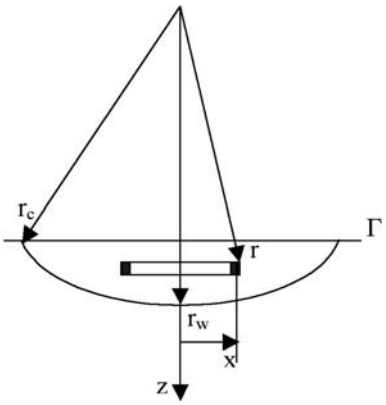


Figure 5-4. Schematic for calculating the boundary force to compensate for a nonperiodic boundary condition

$$F_b(r_w) = -2\pi\rho_n \int_{z=r_w}^{r_c} \int_{x=0}^{\sqrt{r_c^2-z^2}} g(r) \frac{\partial U_{12-6}(r)}{\partial r} \frac{z}{r} x dx dz,$$

and

$$U_b(r_w) = 2\pi\rho_n \int_{z=r_w}^{r_c} \int_{x=0}^{\sqrt{r_c^2-z^2}} g(r) U_{12-6}(r) x dx dz, \tag{5-3}$$

r_c denotes the cutoff radius, ρ_n the average number density, $r = \sqrt{x^2+z^2}$ and r_w the distance to the wall. $F_b(r)$ and $U_b(r)$ for a Lennard–Jones fluid can be obtained by either using a readily available parameterization of $g(r)$ [85] or by performing the integration for Eq. (5-3).

This boundary force does not, however, guarantee that particles will not exit the boundary of the atomistic domain. Particles should, in fact, be able to exit the domain because of the interaction of the atomistic region with the surrounding medium. Therefore, it is necessary to employ a hard specular wall that moves with the local fluid velocity. Particles that strike the wall are specularly reflected, i.e., their velocity components normal to the wall are reversed while the other components are not altered by the impact. A specular wall acts as a plane of symmetry and prevents density perturbations close to it. The instantaneous momentum of the atomistic system is altered by each particle reflection although the average momentum remains constant. At the end of each time step these walls are reset to their initial positions to maintain a fixed frame of reference. The particles remaining outside the computation domain after wall resetting are reinserted in inflow regions following the Usher algorithm [82], which in turn maintains a constant

number of particles for a steady incompressible flow and minimizes disturbances in the potential energy of the system.

MD velocities can be sampled in cells of the same size as the cells in the LBM domain. These velocities can be imposed as boundary conditions for the continuum calculation using two coupling methods. The first [68] amounts to imposing MD velocities within a one cell wide strip located at a distance δ_s from the MD sub domain as shown in Figure 5-3 while the other couples velocity gradients [64].

5.4. EXAMPLE PROBLEMS

Future investigations should examine the effects of various surface treatments on both solid-solid and solid-liquid interface thermal conductances, e.g. through the fundamentals of hydrophobic and hydrophilic surfaces [21, 22], and an examination of the effects of monolayers that can be experimentally self-assembled and surfactants. Molecular dynamics simulations [86–88] predict that the interface thermal conductance should increase significantly with increasing strength of the bonds between the solid and liquid. In particular, the most complete work predicts a substantial difference in conductance between wetting and non-wetting interfaces. However, there are no comprehensive or systematic synergistic simulation and experimental investigations of the effects of solid-liquid bonding on interface conductance. Thus, suggestions for future work involve examinations of semiconducting and oxide materials, which represent some of the simplest systems of hydrophobic and hydrophilic surfaces. Many semiconductors, e.g. Si, are hydrophobic, but in the presence of oxygen many of these surfaces oxidize, e.g., to SiO_2, and become hydrophilic. Since the thermal properties of these materials are well known, they make for ideal test cases. After examining these solid hydrophobic and hydrophilic layers, these investigations should continue by examining self-assembled functional groups that are either hydrophobic or hydrophilic. These self-assembled layers have great technical importance as they may be applied to a wide variety of solid materials.

ACKNOWLEDGMENTS

We are grateful for helpful discussions with Dr. Anindya De of GE Research, Bangalore, India.

REFERENCES

1. Swartz, E.T. and R.O. Pohl, *Thermal-boundary resistance*. Reviews of Modern Physics, 1989. **61**(3): 605–668.
2. Darabi, J., *Micro- and nanoscale heat transfer: Challenges and opportunities*. Heat Transfer Engineering, 2002. **23**(2): 1–2.
3. Chen, G., *Particularities of heat conduction in nanostructures*. Journal of Nanoparticle Research, 2000. **2**(2): 199–204.
4. Thome, J.R., *The new frontier in heat transfer: Microscale and nanoscale technologies*. Heat Transfer Engineering, 2006. **27**(9): 1–3.

5. Kapitza, P.L., *The study of heat transfer in helium II*. Journal of Physics, USSR, 1941. **4**: 181.
6. Balasubramanian, G., S. Banerjee, and I.K. Puri, *Unsteady nanoscale thermal transport across a solid-fluid interface*. Journal of Applied Physics, 2008. **104**(6).
7. Pollack, G.L., *Kapitza resistance*. Reviews of Modern Physics, 1969. **41**(1): 48–81.
8. Ferrell, R.A., J.K. Bhattacharjee, and S.I. Mukhin, *Theory for normal state critical Kapitza resistance of He-4*. Low Temperature Physics, 1998. **24**(2): 76–77.
9. Murad, S. and I.K. Puri, *Thermal transport across nanoscale solid-fluid interfaces*. Applied Physics Letters, 2008. **92**: 133105.
10. Murad, S. and I.K. Puri, *Molecular simulation of thermal transport across hydrophilic interfaces*. Chemical Physics Letters, 2008. **467**(1-3): 110–113.
11. Challis, L.J., *Kapitza resistance and acoustic transmission across boundaries at high frequencies*. Journal of Physics C: Solid State Physics, 1974. **7**(3): 481–495.
12. Barrat, J.L. and F. Chiaruttini, *Kapitza resistance at the liquid-solid interface*. Molecular Physics, 2003. **101**(11): 1605–1610.
13. Khare, R., P. Keblinski, and A. Yethiraj, *Molecular dynamics simulations of heat and momentum transfer at a solid-fluid interface: Relationship between thermal and velocity slip*. International Journal of Heat and Mass Transfer, 2006. **49**(19-20): 3401–3407.
14. Wang, X.W., Z.R. Zhong, and J. Xu, *Noncontact thermal characterization of multiwall carbon nanotubes*. Journal of Applied Physics, 2005. **97**(6).
15. Barrat, J.-L. and F. Chiaruttini, *Kapitza resistance at the liquid-solid interface*. Molecular Physics, 2003. **101**: 1605–1610.
16. Oligschleger, C. and J.C. Schon, *Simulation of thermal conductivity and heat transport in solids*. Physical Review B, 1999. **59**(6): 4125–4133.
17. Che, J.W., et al., *Thermal conductivity of diamond and related materials from molecular dynamics simulations*. Journal of Chemical Physics, 2000. **113**(16): 6888–6900.
18. Schwab, K., et al., *Measurement of the quantum of thermal conductance*. Nature, 2000. **404**(6781): 974–977.
19. Huxtable, S.T., et al., *Interfacial heat flow in carbon nanotube suspensions*. Nature Materials, 2003. **2**: 731–734.
20. Koplik, J., J.R. Banavar, and J.F. Willemsen, *Molecular-dynamics of fluid-flow at solid-surfaces*. Physics of Fluids A-Fluid Dynamics, 1989. **1**(5): 781–794.
21. Murad, S. and I.K. Puri, *Dynamics of nanoscale jet formation and impingement on flat surfaces*. Physics of Fluids, 2007. **19**: 128102.
22. Xue, L., et al., *Two regimes of thermal resistance at a liquid-solid interface*. Journal of Chemical Physics, 2003. **118**(1): 337–339.
23. Challis, L.J., K. Dransfeld, and J. Wilks, *Heat transfer between solids and liquid helium II*. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 1961. **260**(1300): 31–46.
24. Israelachvili, J., *Intermolecular and Surface Forces*, 2nd ed, 1992. London: Academic Press.
25. Naha, S., S. Sen, and I.K. Puri, *Flame synthesis of superhydrophobic amorphous carbon surfaces*. Carbon, 2007. **45**(8): 1702–1706.
26. Sen, S. and I.K. Puri, *Flame synthesis of carbon nanofibres and nanofibre composites containing encapsulated metal particles*. Nanotechnology, 2004. **15**(3): 264–268.
27. Arana, C.P., I.K. Puri, and S. Sen, *Catalyst influence on the flame synthesis of aligned carbon nanotubes and nanofibers*. Proceedings of the Combustion Institute, 2005. **30**: 2553–2560.
28. Ramadurai, K., *Carbon nanostructures for thermal applications: synthesis and characterization*, in *Mechanical Engineering*, 2007. Boulder, CO: University of Colorado.
29. Huhtala, M., A. Kuronen, and K. Kaski, *Carbon nanotube structures: molecular dynamics simulation at realistic limit*. Computer Physics Communications, 2002. **146**(1): 30–37.

30. Rafii-Tabar, H., *Computational modelling of thermo-mechanical and transport properties of carbon nanotubes (vol 390, pg 235, 2004)*. Physics Reports-Review Section of Physics Letters, 2004. **394**(6): 357.
31. Louchev, O.A., et al., *Thermal physics in carbon nanotube growth kinetics*. Journal of Chemical Physics, 2004. **121**(1): 446–456.
32. Yang, X.S., *Modelling heat transfer of carbon nanotubes*. Modelling and Simulation in Materials Science and Engineering, 2005. **13**(6): 893–902.
33. Kotsalis, E.M., J.H. Walther, and P. Koumoutsakos, *Multiphase water flow inside carbon nanotubes*. International Journal of Multiphase Flow, 2004. **30**(7-8): 995–1010.
34. Liu, Y.C., et al., *Fluid structure and transport properties of water inside carbon nanotubes*. Journal of Chemical Physics, 2005. **123**(23).
35. Hanasaki, I. and A. Nakatani, *Water flow through carbon nanotube junctions as molecular convergent nozzles*. Nanotechnology, 2006. **17**(11): 2794–2804.
36. Banerjee, S., S. Murad, and I.K. Puri, *Hydrogen storage in carbon nanostructures: Possibilities and challenges for fundamental molecular simulations*. Proceedings of the IEEE, 2006. **94**(10): 1806–1814.
37. Bolton, K. and S. Gustavsson, *Energy transfer mechanisms in gas-carbon nanotube collisions*. Chemical Physics, 2003. **291**(2): 161–170.
38. Hu, M., et al., *Thermal energy exchange between carbon nanotube and air*. Applied Physics Letters, 2007. **90**(23).
39. Zhong, H.L. and J.R. Lukes, *Interfacial thermal resistance between carbon nanotubes: Molecular dynamics simulations and analytical thermal modeling*. Physical Review B, 2006. **74**(12).
40. Shenogin, S., et al., *Role of thermal boundary resistance on the heat flow in carbon-nanotube composites*. Journal of Applied Physics, 2004. **95**(12): 8136–8144.
41. Shenogin, S., et al., *Effect of chemical functionalization on thermal transport of carbon nanotube composites*. Applied Physics Letters, 2004. **85**(12): 2229–2231.
42. Keblinski, P., et al., *Mechanisms of heat flow in suspensions of nano-sized particles (nanofluids)*. International Journal of Heat and Mass Transfer, 2002. **45**(4): 855–863.
43. Huxtable, S.T., et al., *Interfacial heat flow in carbon nanotube suspensions*. Nature Materials, 2003. **2**(11): 731–734.
44. Vladkov, M. and J.L. Barrat, *Modeling transient absorption and thermal conductivity in a simple nanofluid*. Nano Letters, 2006. **6**(6): 1224–1228.
45. Xue, Y.Q. and M.D. Chen, *Dynamics of molecules translocating through carbon nanotubes as nanofluidic channels*. Nanotechnology, 2006. **17**(20): 5216–5223.
46. Banerjee, S., S. Murad, and I.K. Puri, *Preferential ion and water intake using charged carbon nanotubes*. Chemical Physics Letters, 2007. **434**(4-6): 292–296.
47. Koplik, J. and J.R. Banavar, *Continuum deductions from molecular hydrodynamics*. Annual Review of Fluid Dynamics, 1995. **27**: 257–292.
48. Murad, S. and I.K. Puri, *Nanoscale jet collision and mixing dynamics*. Nano Letters, 2007. **7**(3): 707–712.
49. Choi, Y.S., S.J. Kim, and M.U. Kim, *Molecular dynamics of unstable motions and capillary instability in liquid nanojets*. Physical Review E, 2006. **73**(1), Art. No. 016309.
50. Fang, T.H., W.J. Chang, and S.C. Liao, *Simulated nanojet ejection process by spreading droplets on a solid surface*. Journal of Physics: Condensed Matter, 2003. **15**(49): 8263–8270.

51. Moseler, M. and U. Landman, *Formation, stability, and breakup of nanojets*. Science, 2000. **289**(5482): 1165–1169.
52. Heyes, D.M. and N.H. March, *Theoretical approaches to thermal conductivity in liquids*. Physics and Chemistry of Liquids, 1996. **33**(2): 65–83.
53. Eastman, J.A., et al., *Thermal transport in nanofluids*. Annual Review of Materials Research, 2004. **34**: 219–246.
54. Keblinski, P., et al., *Thermodynamic behavior of a model covalent material described by the environment-dependent interatomic potential*. Physical Review B, 2002. **66**(6).
55. Bodapati, A., et al., *Vibrations and thermal transport in nanocrystalline silicon*. Physical Review B, 2006. **74**(24).
56. Tang, Q.H. and Y.G. Yao, *The Kapitza resistance across grain boundary by molecular dynamics simulation*. Nanoscale and Microscale Thermophysical Engineering, 2006. **10**(4): 387–398.
57. Schelling, P.K., S.R. Phillpot, and P. Keblinski, *Kapitza conductance and phonon scattering at grain boundaries by simulation*. Journal of Applied Physics, 2004. **95**(11): 6082–6091.
58. Stevens, R.J., L.V. Zhigilei, and P.M. Norris, *Effects of temperature and disorder on thermal boundary conductance at solid-solid interfaces: Nonequilibrium molecular dynamics simulations*. International Journal of Heat and Mass Transfer, 2007. **50**(19-20): 3977–3989.
59. Patel, H.A., S. Garde, and P. Keblinski, *Thermal resistance of nanoscopic liquid-liquid interfaces: Dependence on chemistry and molecular architecture*. Nano Letters, 2005. **5**(11): 2225–2231.
60. Twu, C.J. and J.R. Ho, *Molecular-dynamics study of energy flow and the Kapitza conductance across an interface with imperfection formed by two dielectric thin films*. Physical Review B, 2003. **67**(20).
61. Angadi, M.A., et al., *Thermal transport and grain boundary conductance in ultrananocrystalline diamond thin films*. Journal of Applied Physics, 2006. **99**(11).
62. Hegedus, P.J. and A.R. Abramson, *A molecular dynamics study of interfacial thermal transport in heterogeneous systems*. International Journal of Heat and Mass Transfer, 2006. **49**(25-26): 4921–4931.
63. Chiritescu, C., et al., *Ultralow thermal conductivity in disordered, layered WSe2 crystals*. Science, 2007. **315**(5810): 351–353.
64. Dupuis, A., E.M. Kotsalis, and P. Koumoutsakos, *Coupling lattice Boltzmann and molecular dynamics models for dense fluids*. Physical Review E, 2007. **75**(4).
65. Xiao, S.P. and T. Belytschko, *A bridging domain method for coupling continua with molecular dynamics*. Computer Methods in Applied Mechanics and Engineering, 2004. **193**(17-20): 1645–1669.
66. Xu, J.L. and Y.X. Li, *Boundary conditions at the solid-liquid surface over the multiscale channel size from nanometer to micron*. International Journal of Heat and Mass Transfer, 2007. **50**(13–14): 2571–2581.
67. Nie, X.B., S.Y. Chen, and M.O. Robbins, *Hybrid continuum-atomistic simulation of singular corner flow*. Physics of Fluids, 2004. **16**(10): 3579–3591.
68. Werder, T., J.H. Walther, and P. Koumoutsakos, *Hybrid atomistic-continuum method for the simulation of dense fluid flows*. Journal of Computational Physics, 2005. **205**(1): 373–390.
69. Cahill, D.G., et al., *Nanoscale thermal transport*. Journal of Applied Physics, 2003. **93**(2): 793–818.
70. Khater, A.F., *The Kapitza resistance and phonon scattering at solid-liquid He interfaces*. Le Journal de Physique Colloques, 1978. **39**(C6).
71. Chen, S. and G.D. Doolen, *Lattice Boltzmann method for fluid flows*. Annual Review of Fluid Mechanics, 1998. **30**: 329–364.
72. Succi, S., *Lattice Boltzmann equation: Failure or success?*. Physica A, 1997. **240**(1–2): 221–228.
73. De, A.K., A. Mukhopadhyay, and I.K. Puri, *Lattice Boltzmann method simulation of electroosmotic stirring in a microscale cavity*. Microfluidics and Nanofluidics, 2008. **4**(5): 463–470.

74. Succi, S., *The Lattice Boltzmann Equation for Fluid Dynamics and Beyond*, 2001. New York: Oxford.
75. Nourgaliev, R.R., et al., *The lattice Boltzmann equation method: theoretical interpretation, numerics and implications*. International Journal of Multiphase Flow, 2003. **29**(1): 117–169.
76. Bhatnagar, P.L., E.P. Gross, and M. Krook, *A model for collisional processes in gases I: Small amplitude processes in charged and neutral one-component system*. Physical Review, 1954. **94**: 511.
77. He, X., S. Chen, and G.D. Doolen, *A novel thermal model for the lattice Boltzmann method in incompressible limit*. Journal of Computational Physics, 1998. **146**(1): 282–300.
78. Zou, Q.S. and X.Y. He, *On pressure and velocity boundary conditions for the lattice Boltzmann BGK model*. Physics of Fluids, 1997. **9**(6): 1591–1598.
79. O'Connell, S.T. and P.A. Thompson, *Molecular dynamics-continuum hybrid computations: A tool for studying complex fluid flows*. Physical Review E, 1995. **52**(6): R5792–R5795.
80. Sun, Q.H., I.D. Boyd, and G.V. Candler, *A hybrid continuum/particle approach for modeling subsonic, rarefield gas flow*. Journal of Computational Physics, 2004. **194**(1): 256–277.
81. Flekkoy, E.G., G. Wagner, and J. Feder, *Hybrid model for combined particle and continuum dynamics*. Europhysics Letters, 2000. **52**(3): 271–276.
82. Delgado-Buscalioni, R. and P.V. Coveney, *USHER: An algorithm for particle insertion in dense fluids*. Journal of Chemical Physics, 2003. **119**(2): 978–987.
83. Hadjiconstantinou, N.G. and A.T. Patera, *Heterogeneous atomistic-continuum representations for dense fluid systems*. International Journal of Modern Physics C, 1997. **8**(4): 967–976.
84. Hadjiconstantinou, N.G., *Hybrid atomistic-continuum formulations and the moving contact-line problem*. Journal of Computational Physics, 1999. **154**(2): 245–265.
85. Matteoli, E. and G.A. Mansoori, *A simple expression for radial-distribution functions of pure fluids and mixtures*. Journal of Chemical Physics, 1995. **103**(11): 4672–4677.
86. J.-L. Barrat and F. Chiaruttini, *Kapitza resistance at the liquid-solid interface*. Molecular Physics, 2003. **101**: 1605–1610.
87. S. Maruyama and T. Kimura, *A study on thermal resistance over a solid-liquid interface by the molecular dynamics method*. Thermal Science Engineering, 1999. **7**: 63–68.
88. L. Xue, et al., *Two regimes of thermal resistance at a liquid-solid interface*. Journal of Chemical Physics, 2003. **118**: 337–339.

CHAPTER 6

MULTISCALE MODELING OF CONTACT-INDUCED PLASTICITY IN NANOCRYSTALLINE METALS

VIRGINIE DUPONT[1] AND FREDERIC SANSOZ[2]

[1]*School of Engineering, University of Vermont, Burlington, VT 05405, USA, e-mail: virginiedupont@gmail.com*

[2]*School of Engineering and Materials Science Program, University of Vermont, Burlington, VT 05405, USA,e-mail: frederic.sansoz@uvm.edu*

Abstract: Predicting the integrity of metallic thin films deposited on semiconductors for micro-electromechanical systems (MEMS) applications requires a precise understanding of surface effects on plasticity in materials with nano-sized grains. Experimentally, the use of nanoscale contact probes has been very successful to characterize the dependence of flow stress on mean grain size in nanocrystalline metals. From atomistic simulations, several models of plastic yielding for metal indentation have also been proposed based on the nucleation and propagation of lattice dislocations, and their interaction with grain boundaries beneath penetrating tips. However, model refinement is needed to include the characteristics of materials whose grain size is much smaller than the typical plastic zones found in contact experiments. Particularly, cooperative deformation processes mediated by grain boundaries, such as grain rotation, deformation twinning, and stress-driven grain coarsening, can simultaneously emerge for very small grain sizes (< 20 nm), thus making a predictive understanding of plastic yielding elusive. This chapter summarizes our recent progress in using multiscale modeling to gain fundamental insight into the underlying mechanisms of surface plasticity in nanocrystalline face-centered cubic metals deformed by nanoscale contact probes. Two numerical approaches to model contact-induced plasticity in nanocrystalline materials, the quasicontinuum method and parallel molecular dynamics simulation, are reviewed. Using these techniques, we discuss the role of a grain boundary network on the incipient plasticity of nanocrystalline Al films deformed by wedge-like cylindrical tips, as well as the processes of stress-driven grain growth in nanocrystalline films subjected to nanoindentation

Keywords: Nanoindentation, Nanocrystalline metal, Atomistic simulation

6.1. INTRODUCTION

Nanocrystalline films of pure face-centered cubic (FCC) metals such as Al, Ni and Cu with a grain size less than 100 nm, are commonly used in surface-micromachining to process micron-scale structures and devices including micro-electromechanical systems (MEMS) [1–5]. Characterizing the mechanical and

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 151–172.
DOI 10.1007/978-1-4020-9785-0_6,

tribological behavior of nanocrystalline metals upon contact loading remains an essential task in predicting the structural integrity of such micro-components [6, 7]. In addition, past studies [8–10] have proved that a marked transition in plasticity mechanisms operates with a reduction of grain size from the microcrystalline to the nanocrystalline regime in FCC metals.

While advanced characterization techniques have been used to study the mechanical behavior of thin films, such as MEMS-based tensile testing [11–14] or membrane deflection experiments [15], small-scale contact experiments, such as nanoindentation, have been used ubiquitously to characterize the nature of yield phenomena and the influence of grain size on hardness and strengthening in nanocrystalline metals [16–22]. A nanoscale tip attached to the cantilever beam of an atomic force microscope (AFM) that was made to measure the nanomechanical behavior of metallic thin films is shown in Figure 6-1 along with nanoscale indentations performed on a nanocrystalline Ni surface. Such probes are particularly well-suited for the studies of plasticity transition in nanograined metals, because they can be highly sensitive to the heterogeneous nature of plastic deformation in very confined volumes of materials.

In earlier studies [9, 24, 25], particular focus has been placed on examining how dislocations interact with surrounding grain boundaries (GBs) by performing nano-indentations at the center of single nanograins, that is, by forcing the contact area to be much smaller than the grain size. Yang and Vehoff [25] have observed that the dislocations, which nucleate below the tip, only interact directly with the neighboring interfaces for grain sizes below 900 nm. At this scale, the point of elastic instability is clearly defined by a pop-in event whose width is strongly correlated to the size of the indented grain. The smaller the grain size, the smaller the pop-in width and the harder the material. For grain sizes comparable to the contact area, Minor et al. [24] have also revealed using in situ transmission electron microscopy (TEM) nanoindentation that significant dislocation activity could take

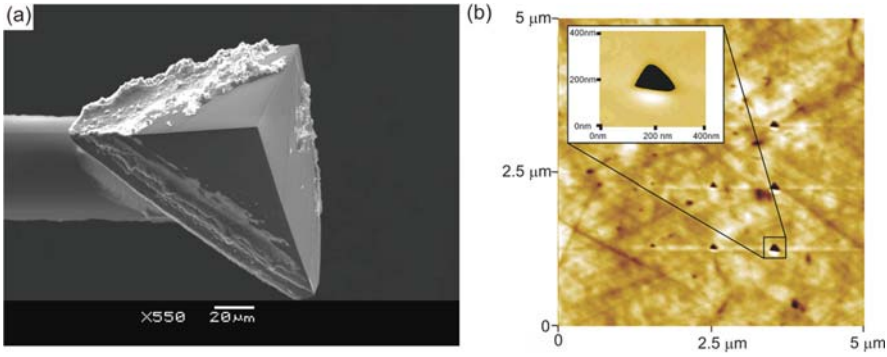


Figure 6-1. Measurements of nanomechanical properties in thin films using nanoscale contact probes. (**a**) SEM image of a made-to-measure cube-corner tip attached to the cantilever beam of an AFM; (**b**) AFM-enabled nanoscale indentations performed on a 50-μm-thick nanocrystalline Ni film [23]

place in ultrafine-grained Al thin films before the first obvious jump in displacement in the load-depth nanoindentation curves. However, a predictive understanding of plastic yielding in nanocrystalline metals during indentation remains elusive, primarily for two reasons:

- While the nanoscale indentations shown in Figure 6-1b can be considered as physically small, these contact zones can still be an order of magnitude larger than the mean grain size of the film tested. In this case, it is the cooperative deformation of the dense GB network that dominates the plastic behavior in nanocrystalline materials. All experimental evidence shows that the pile-up of deformation left around residual impressions varies dramatically from homogeneous at large grain size (> 20 nm) to inhomogeneous with intense plastic deformation in highly-localized shear bands for very small grain sizes (< 20 nm) [9, 21, 26, 27]. Furthermore, new deformation mechanisms, including grain rotation, deformation twinning, and stress-driven grain coarsening, can simultaneously emerge for very small grain sizes [28–32]. Therefore, some model refinement is required in order to include the characteristics of materials whose grain size is much smaller than the typical plastic zones found in contact experiments.
- GB-mediated deformation phenomena have been observed during the nanoindentation of nanocrystalline FCC metals. Abnormal grain growth was found by Jin et al. [29] during in situ TEM experiments of nanoindentation in nanocrystalline Al films. For instance, Figure 6-2 shows different snapshots from Jin et al.'s study, where a local increase in brightness related to grain rotation, indicates that grain growth takes place under the tip as indentation proceeds. Stress-driven grain growth was also observed by Zhang et al. [30, 31] during the micro-indentation of nanocrystalline Cu films. These authors showed that grain growth occurred at faster rates at cryogenic temperature than at room temperature, and that the purity of the material influences the grain growth process. Since it is well-established that the flow stress characterized by hardness testing strongly depends on the material grain size [8, 33–35], the process of contact-induced grain growth appears to be undesirable for most microscale systems with nanocrystalline characteristics. It is therefore critically important to understand the mechanisms of GB motion under nanoscale contact in order to achieve control over such phenomena.

This chapter presents an overview of quasicontinuum (QC) method and large-scale molecular dynamics (MD) simulations used to shed light on the fundamental mechanisms of surface plasticity in pure nanocrystalline metals deformed by nanoscale tips. The numerical methods related to the atomistic modeling of nanoscale contact in nanocrystalline films are reviewed in Section 6.2. Section 6.3 describes the effects of interatomic potentials on randomly-oriented microstructures and the energies at GBs predicted by atomistic simulation. Using the QC method, we focus in Section 6.4 on the role of a GB network on the incipient plasticity of a 7 nm-grain-size nanocrystalline Al film deformed by a wedge-like cylindrical tip. In

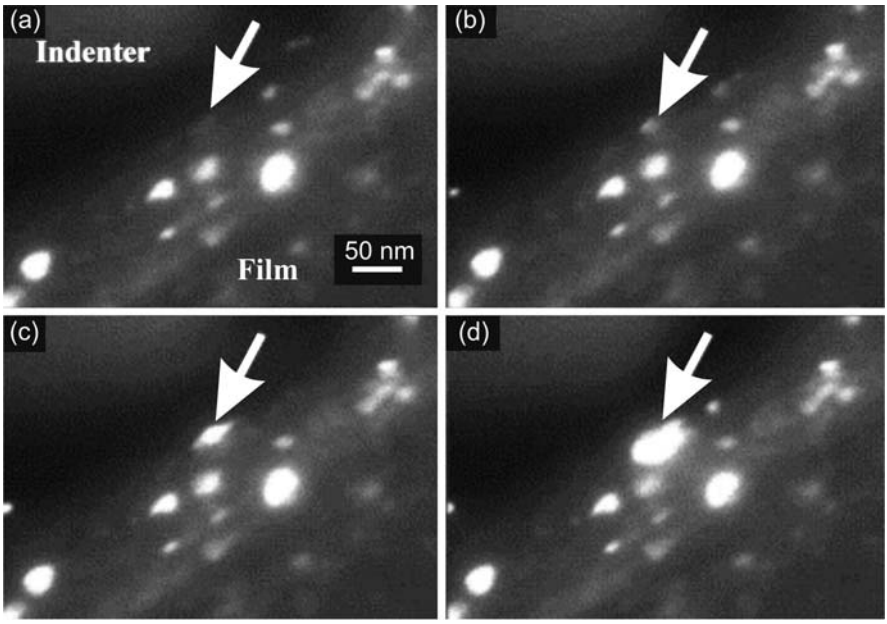


Figure 6-2. Snapshots during in situ TEM nanoindentation of a polycrystalline Al film showing abnormal growth of nano-sized grains. (**a**) No grains in strong diffraction condition under the tip area indicated by the white arrow. (**b**) A grain with size about 10 nm has appeared. (**c**)–(**d**) The size of the grain has become larger with increasing load. Reprinted from [29] with permission from Elsevier

Section 6.5, using both MD and QC methods, we discuss the mechanisms of stress-driven grain growth in nanocrystalline Al during nanoindentation. An outlook for future research in this rich area is also presented in Section 6.6.

6.2. ATOMISTIC MODELING OF NANOSCALE CONTACT IN NANOCRYSTALLINE FILMS

Earlier attempts made to model the nanoindentation of nanocrystalline materials by classical MD simulations have usually employed a spherical repulsive force to model virtual tips varying from 30 to 98 Å in diameter [36–42]. As such, contact areas were, to a large extent, smaller than the grain size, and the plastic zone produced by these tips was only limited to one or two grains. In contrast, Szulfarska et al. [43] have simulated the nanoindentation of normally-brittle nanocrystalline ceramics with a four to one ratio between tip diameter and grain size, which revealed unusual GB-mediated plastic behavior. Different numerical approaches have also been used to study the problems of contact in single-crystal and nanocrystalline films using concurrent multiscale modeling [44–48]. Using the QC method, we show here that it is critically important to simulate nanoindentation tips with more realistic sizes. Furthermore,

atomistic simulations in FCC metals have been performed using different embedded-atom-method (EAM) potentials, from which predictions of stacking fault energies can lead to strong differences within the same metal [49]. The impact of the interatomic potential on cooperative plastic processes, however, is not fully understood. This chapter focuses on the modeling of nanoscale contact in nanocrystalline metal films using MD simulations with LAMMPS [50] and the QC simulation technique [51]. These two methods and specific numerical tools for the analysis of stresses and the visualization of defects and GBs are briefly recalled in the following.

6.2.1. Simulation Methods

6.2.1.1. Molecular Dynamics

Classical MD simulation is a common technique for the numerical investigation of physical and dynamical properties of matter at the molecular level. Each atom in the simulation is treated as a point mass whose velocity and position are computed by time integration of the Newton's equations. The computational task in a typical MD simulation is to solve the set of coupled differential equations given by

$$m_i \frac{d\vec{v}_i}{dt} = \sum_j F_2\left(\vec{r}_i,\vec{r}_j\right) + \sum_j \sum_k F_3\left(\vec{r}_i,\vec{r}_j,\vec{r}_k\right) + \ldots \tag{6-1}$$

$$\frac{d\vec{r}_i}{dt} = \vec{v}_i, \tag{6-2}$$

where m_i is the mass of atom i, $\vec{r}_i$ and $\vec{v}_i$ are its position and velocity vectors, respectively, F_2 is a force function describing pairwise interactions between atoms, F_3 describes three-body interactions, and many-body interactions can be added in the same way. Numerical integration of the atomic positions is usually performed using the Verlet method [52] with a time step, which may be varied from 1 to 5 fs for studies in crystal plasticity. The calculations are also conducted at constant temperature (NVT) using a Nose/Hoover temperature thermostat [53]. Furthermore, the calculations on large-scale systems must take full advantage of massively-parallel computing with open-source MD simulation software such as LAMMPS [50].

6.2.1.2. Quasicontinuum (QC) Method

The QC method, which was developed by Miller and Tadmor [51], is a multiscale atomistic/continuum simulation technique combining both finite element and molecular statics methods. This technique therefore alleviates the need for representing all the atoms as in classical MD simulation. A complete description of the QC method can be found in the review by Miller and Tadmor [51]. A typical QC model consists of atomistic zones (non local) and finite element zones (local). The regions subjected to high plastic deformations are modeled atomically, while the rest is modeled by finite elements. Each node in the mesh is a representative atom, or "repatom", which can either represent just itself (non local zone as well as some

atoms of the atomistic/continuum interface), or a group of atoms (local zone). The total energy of the system is therefore computed such as

$$E_{\text{tot}} = \sum_{\alpha=1}^{N_{\text{rep}}} n_\alpha E_\alpha \approx E_{\text{exact}}, \tag{6-3}$$

where N_{rep} is the total number of repatoms in the system, n_α is the number of real atoms the repatom is representing ($n_\alpha = 1$ for non local atoms), and E_α is the energy of each repatom. This formulation permits to conduct concurrent calculations on both local and non-local regions, which largely decreases the number of degrees of freedom as opposed to MD simulations with corresponding model dimensions. Some discontinuities in energy may also appear at the local/non-local interface, but these can be eliminated by taking ghost forces [51] into account. The minimum energy is calculated at each loading step, after a new set of forces or displacements is applied, using a conjugate gradient method. The QC method can also apply a "nonlocality criterion" to the model in order to verify whether atoms should be local or non-local. This criterion is defined by:

$$\max_{a,b;k} \left| \lambda_k^a - \lambda_k^b \right| < \varepsilon, \tag{6-4}$$

where λ_k^a is the kth eigenvalue of the right stretch tensor $U_a = \sqrt{F_a^T F_a}$ obtained from the deformation gradient F_a in element a, $k = 1\ldots3$, and the indices a and b run over all elements within a cutoff distance of a given repatom. The threshold ε is usually determined empirically, but a value of 10% gives reasonable results in crystal plasticity. If the inequality is not satisfied, the repatom is made non local and vice-versa. It is worth noting that, because of the energy minimization procedure, the QC method can only predict the athermal behavior of crystalline materials at 0 K. Furthermore, Eq. (6-4) may be used with remeshing techniques in order to adapt the finite element mesh as a function of the deformation. Some examples of adaptive QC simulations can be found elsewhere [54, 55].

6.2.2. Modeling of Spherical/Cylindrical Contact in Nanocrystalline Metals

Two methods have been used to model tips in contact problems using atomistic simulation. The first method consists in assuming the tip to be either spherical or cylindrical with a virtual repulsive force such that:

$$F(r) = -k(r - R)^2 \tag{6-5}$$

with k a specified force constant ($k = 10$ N/m^2), R the tip radius, and r the distance between the atom and the center of the tip. This method removes the effects of adhesion and friction that are typically applied by real indenters. Second, the tip can

be modeled by representing all the atoms. If kept rigid, solid tips can be used for direct comparison with different contact theories including adhesion effects [56].

A Voronoi tessellation based on a constrained-Delaunay connectivity scheme [57] is commonly used to model polycrystalline films with randomly-oriented microstructures, i.e. similar to natural GB networks. Reference points are first randomly placed at a specified mean distance from each other in the surface or volume studied. Each reference point becomes the center of a grain whose crystallographic orientation is also randomly assigned. The Voronoi tessellation created from these points forms the network of GBs that are orthogonal to the lines joining the reference point to neighbor reference points as illustrated in Figure 6-3.

In Figure 6-4, we present the use of this methodology to model the indentation of a three-dimensional polycrystalline thin film with a mean grain size of 7 nm. The total number of atoms for such a model is ∼3 million. The film thickness is 30 nm. The film is indented by displacing a spherical tip with a diameter of 18 nm along the direction normal to the top surface. The bottom of the film is fixed along the direction of indentation, while the sides of the model are assigned periodic boundary conditions. The model is first relaxed under zero force condition using a conjugate gradient method in order to obtain the lowest state of energy in the GB network to simulate equilibrium conditions. After relaxation, the tip is moved at a speed on the order of 1 m**.** s^{-1}. The simulation is performed at 300 K with a time step of 5 fs and the atomic positions are recorded at 50 ps intervals (10,000 steps).

Similarly, a QC model of indentation in a 200-nm-thick Al film with a wedge-like cylindrical tip is shown Figure 6-5. In this model, the contact region at the interface between the indenter and the film surface is fully represented by individual atoms. For comparison, the dimensions of both film and fully-atomistic zone are $400 \times 200 \times 0.286$ nm and $50 \times 25 \times 0.286$ nm, respectively. Plane-strain contact is modeled by displacing a single-crystal Al cylinder with a diameter of 30 nm along the direction normal to the film. The tip is oriented along the crystallographic

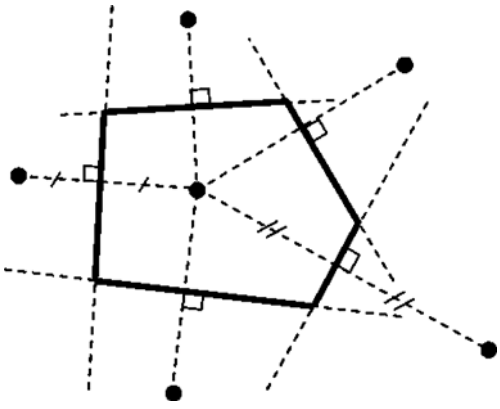


Figure 6-3. Schematic showing the Voronoi construction for a two-dimensional polycrystalline model

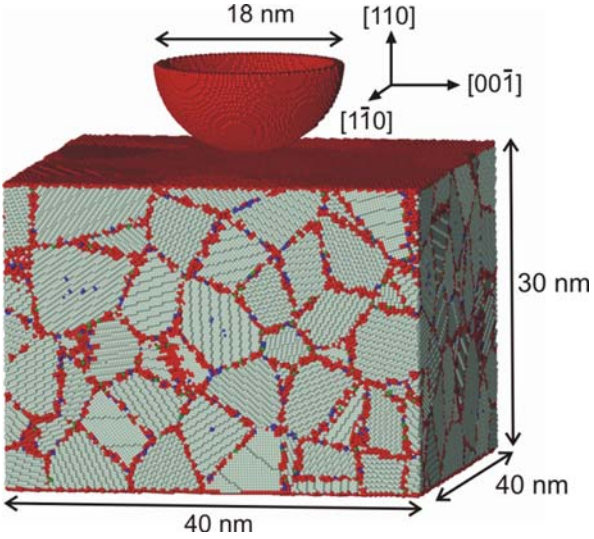


Figure 6-4. Molecular dynamics model of a nanocrystalline thin film to be indented by a rigid, spherical tip. Tip diameter and mean grain size are 18 and 7 nm, respectively. The crystallographic orientations of the tip are as indicated. Periodic boundary conditions are assigned on each side. This model consists of ~3 million atoms

directions shown in Figure 6-5b and kept completely rigid during the simulation. This model consists of ~40,000 atoms.

To avoid discontinuities in the energy state during force minimization, the continuum/atomistic frontier is modeled as a single crystal interface with the crystallographic orientations shown in Figure 6-5a. We note that if the deformation is small (tip penetration < 10 nm), no significant atomistic activity is found near this interface, indicating that the plastic deformation is limited to the polycrystalline region during the simulations. The bottom of the film is fixed along each direction, while both sides of the model are left free. Periodic boundary conditions are imposed along the out-of-plane direction in the entire model. Each grain is assigned a common tilt axis along the [110] direction, and random in-plane orientation. Therefore, this QC model simulates the mechanical behavior of a randomly oriented two-dimensional columnar microstructure, which may differ from that of fully three-dimensional polycrystalline microstructures. Similar to the MD model, the sample is first relaxed under zero pressure condition in order to obtain the equilibrium microstructure. After relaxation, energy minimization by conjugate gradient method is performed between each loading step.

6.2.3. Calculations of Local Stresses and Mean Contact Pressures

In the fully-atomistic zone of QC models, the local stress tensor of the *i*th atom, σ_i, is calculated using the formula provided by Lilleodden et al. [38], which can be simplified as follows:

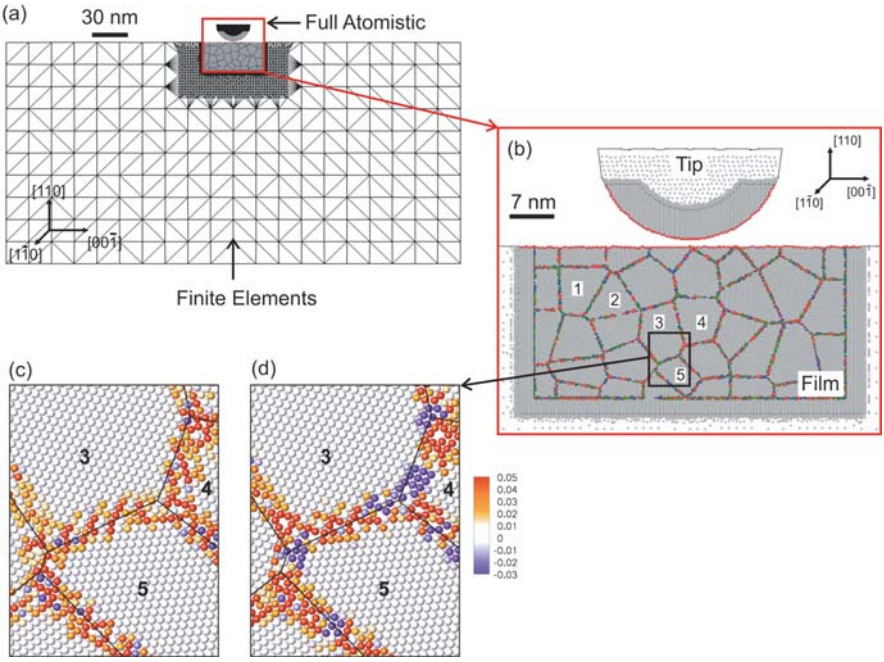


Figure 6-5. Quasicontinuum model of a 7 nm-grain-size Al film indented by a 30-nm-diameter cylindrical tip. (**a**) Full view of both finite element and fully-atomistic regions. (**b**) Close-up view of the full atomistic zone near the contact region. (**c**) Structure and atomic energies at grain boundaries between three nanograins as predicted from the Voter-Chen potential, and (**d**) the Mishin-Farkas potential. Adapted from reference [44] with permission by Elsevier

$$\sigma_i^{\alpha\beta} = \frac{1}{\omega_i^0 \mathrm{Det}[F_i^{\alpha\beta}]} \left\{ \frac{1}{2} \sum_j \left(-\frac{1}{r} \frac{\partial \varphi}{\partial r} \right) r^\alpha r^\beta \Big|_{r=r_{ij}} \right\}, \tag{6-6}$$

where α and β are the Cartesian coordinates, ω_i^0 is the undeformed atomic volume of atom i, $\mathrm{Det}[F_i{}^{\alpha\beta}]$ is the determinant of the deformation gradient, ϕ is the interatomic potential, and r_{ij} is the distance between ith and jth atoms. Note that the kinetics terms have been eliminated in Eq. (6-6) as compared to the formula in [38]. In this equation, the use of the determinant of the deformation gradient has been shown to provide improved accuracy for the calculation of deformed atomic volumes. Furthermore, the force applied by solid tips is calculated using

$$F = \sum_{i \in Z} F_i, \tag{6-7}$$

where Z represents all atoms of the contact area belonging to the film and F_i is the out-of-balance force on atom i in this area, projected along the direction of indentation. The contact area is computed after each loading step by only including atoms at the indenter-film interface within a separation distance from the tip equal to the potential cutoff radius. For wedge-like, cylindrical tips, the mean contact pressure H can be determined at each step such as

$$H = \frac{F}{2a \times z_{\text{perio}}}, \tag{6-8}$$

where a is the contact length, defined as half the width of the projected contact area, and z_{perio} is the sample thickness in the out-of-plane direction.

For virtual tips with a repulsive force, if an atom has penetrated the boundary of the indenter (e.g., $R > r$) after the loading step is completed, the atom is considered "contacted" and feels a force given by Eq. (6-5). The force applied to all contacted atoms is resolved in the direction of indentation, summed and recorded, and is used in the generation of load–displacement profiles [38]. The mean contact pressure is calculated by taking the total applied force on the tip, and by dividing this value by the projected contact area calculated from the position of the contacted atoms.

6.2.4. Tools for the Visualization of Defects and Grain Boundaries

6.2.4.1. Centro-Symmetry Parameter

In solid-state systems, the centro-symmetry parameter P is a useful measure of the local lattice disorder around an atom and can be used to characterize whether the atom is part of a perfect lattice, a local defect (e.g., a dislocation or stacking fault) or at a surface. P is computed using the following formula [58]:

$$P = \sum_{i=1,6} \left| \vec{R}_i + \vec{R}_{i+6} \right|^2, \tag{6-9}$$

where the 12 nearest neighbors are found and R_i and R_{i+6} are the vectors from the central atom to the opposite pair of nearest neighbors. An atom in perfect FCC lattice has a centro-symmetry parameter of zero. The values for other configurations depend on the material chosen. For aluminum, those values are 32.8 Å^2 for a surface atom, 8.2 Å^2 for atoms in an intrinsic stacking fault, and 2.05 Å^2 for atoms halfway between fcc and hcp sites (in a partial dislocation). Equation (6-9) is generally projected in the plane for use with the QC method. In two-dimension, these values are 16.4 Å^2, 4.1 Å^2 and 1.025 Å^2, respectively, for the QC method. The centro-symmetry parameter is well adapted for atomistic simulations at low temperatures, since no averaging for P is needed in these conditions. In the following, atoms in a perfect FCC lattice are either colored in grey or omitted for clarity, those

with a HCP structure or representing a stacking fault are in blue color, and all other non-coordinated atoms appear in green or red colors.

6.2.4.2. Local Crystal Structure by Ackland and Jones

In contrast to the centro-symmetry parameter, the method using the formulation by Ackland and Jones [59] averages out statistical fluctuations due to the temperature boost, because it is not based on the distance between atoms, but the angles between atom pairs. This parameter classifies atoms depending on the closest crystallographic structure it belongs to (BCC, FCC, HCP or unknown). The procedure [59] first calculates the mean squared separation

$$r_0^2 = \sum_{j=1,6} r_{ij}^2 \Big/ 6, \tag{6-10}$$

for the nearest six particles to atom i (i.e., the closest neighbors that verify $r_{ij}^2 < 1.55 r_0^2$). For each of the neighbor pairs found, the bond angle cosines $\cos(\theta_{jik})$ is determined. The procedure, which enables determining the local crystal structure to which atom i can be assigned, relies on a table provided by Ackland and Jones that separates the possible cosine values into 8 ranges. The color scheme used for this parameter is the same than that for the centro-symmetry parameter.

6.3. EFFECTS OF INTERATOMIC POTENTIALS ON EQUILIBRIUM MICROSTRUCTURES

Interactions amongst atoms for FCC metals are represented using an embedded-atom-method (EAM) potential [60], which most accurately predicts the energies of defects and surfaces in such metals. The total energy of a monoatomic system by EAM is described by [61]:

$$E_{\text{tot}} = \frac{1}{2} \sum_{ij} V\left(r_{ij}\right) + \sum_{i} F\left(\bar{\rho}_i\right) \tag{6-11}$$

where $V(r_{ij})$ is a pair potential as a function of the distance r_{ij} between atoms i and j, and F is the embedding energy as a function of the host density $\bar{\rho}_i$ induced at site i by all other atoms in the system. The latter is given by:

$$\bar{\rho}_i = \sum_{j \neq i} \rho(r_{ij}), \tag{6-12}$$

$\rho(r_{ij})$ being the atomic density function. The second term in Eq. (6-11) is volume dependent and represents, in an approximate manner, many-body interactions in the system. EAM potentials are fitted to experimental and ab-initio data for the values

of equilibrium lattice parameter, the cohesive energy, the elastic constants and the vacancy formation energy. This basic set of properties can often be complemented by other data such as planar fault energy and phonon frequencies.

Hereafter, we focus our attention on the effect of EAM interatomic potentials in predicting generalized stacking and planar fault energies, as well as GB energies in equilibrium microstructures. To this end, we compare both the Mishin-Farkas [61] and Voter-Chen [62] potentials for Al. For brevity in the following, these two potentials are referred to as Al-VC and Al-MF potentials, respectively. For each potential, past QC procedures [63, 64] were used to calculate the generalized planar and stacking fault energy curves and the GB energy of 18 Σ <110> tilt bicrystals consisting of symmetrical tilt GBs. The generalized stacking and planar fault energy curves for both potentials are shown in Figure 6-6a. The unstable stacking fault energy (γ_{USF}), stacking fault energy (γ_{SF}) and unstable twinning fault energy (γ_{UTF}) are also indicated in this figure. We find that the calculated energy values are significantly smaller for the Al-VC potential than the Al-MF potential, which is consistent with the predicted values in the literature [61]. In addition, we find that all the ratios γ_{SF}/γ_{USF} and $\gamma_{UTF}/\gamma_{USF}$ are similar and equal to 0.81–0.86 and 1.30–1.32, respectively, which suggests the same slip and twinning behavior regardless of the interatomic potential for Al [65]. Figure 6-6b also shows a significant increase in GB energy for the symmetric tilt bicrystals from the Al-VC potential to the Al-MF potential. The difference of GB energy at atomic level, as a function of EAM potential, is also clearly shown in Figures 6-5c and d, where a cluster of three nanograins is represented. Both atomic energies and structures at GBs are dramatically changed by the interatomic potential, despite the same misorientation angles between grains. This observation therefore indicates that the bonding properties of the GBs will be markedly different depending on the potential.

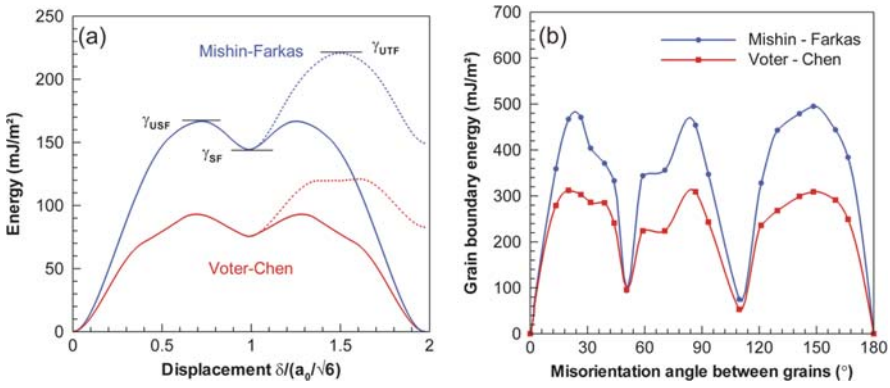


Figure 6-6. Comparison of the Mishin-Farkas and Voter-Chen EAM potentials for Al by QC method for (**a**) the generalized stacking fault energy (*solid line*) and planar fault energy (*dashed line*) curves, and (**b**) GB energies in symmetrical <110>-tilt bicrystals

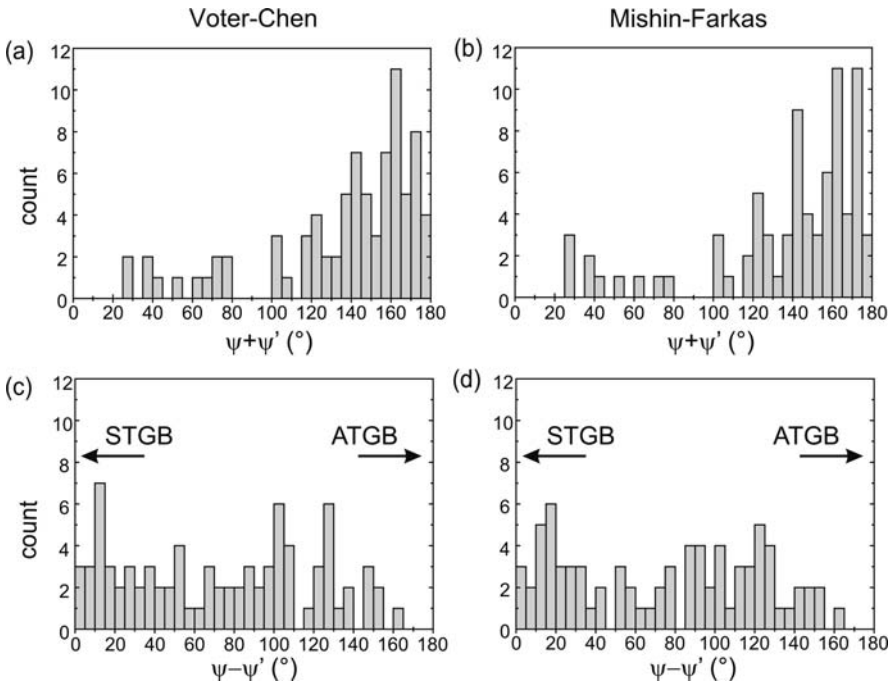


Figure 6-7. Effects of EAM potentials on equilibrium microstructures in the simulated nanocrystalline Al film shown in Figure 6-5. (**a**)–(**b**) Distribution of misorientation angles ($\psi + \psi'$) between grains. (**c**)–(**d**) Degree of symmetry of the GB structure from perfectly-symmetrical tilt GB (STGB, $\psi - \psi' \sim 0$) to highly-asymmetrical tilt GB (ATGB, $\psi - \psi' \sim 180°$). Reprinted from [44] with permission from Elsevier

Furthermore, the effect of potential is investigated on the misorientation angle and asymmetry of the GBs in the 7-nm-grain-size microstructure shown in Figure 6-5. Figure 6-7 shows the angles ψ and ψ', which represent the angle between the [110] lattice direction and the GB plane for two neighboring grains, respectively. The sum $\psi + \psi'$ characterizes the misorientation angle at the boundary separating two grains. Some minor differences in the misorientation angle $\psi + \psi'$ are observed due to the rearrangement of microstructure after relaxation, even though the original model is identical. It can however be concluded from Figures 6-7a and b that the distribution of misorientation angle and, therefore, the overall microstructure of the GB network are not significantly affected by the potential. In addition, the difference $\psi - \psi'$ represents the degree of symmetry of the GB, where $\psi - \psi' = 0°$ corresponds to perfectly-symmetric tilt GBs, and $\psi - \psi' = 180°$ to highly-asymmetric tilt GBs. In Figures 6-7c and d, the degree of symmetry is found to be homogeneous along the set of values for $\psi - \psi'$ regardless of the potential.

In summary, the two EAM potentials for Al mostly differ in their prediction of the generalized stacking and planar fault energies, and GB energies, but should not significantly influence the grain morphology at equilibrium and slip behavior in the material.

6.4. EFFECTS OF A GRAIN BOUNDARY NETWORK ON INCIPIENT PLASTICITY DURING NANOSCALE CONTACT

Shear banding is an important mode of plastic deformation in nanocrystalline materials, and is best illustrated by a 5 nm-grain-size simulation using the Al-VC potential (Figure 6-8). At the onset of plasticity, significant GB sliding takes place leading to rotational deformation of the grains with limited intragranular slip. During this process, the GB structure is significantly changed and, in some cases, several GBs tend to be aligned (Figure 6-8b). The bands are formed by the sliding of aligned interfaces separating the grains (see for example grains 3 and 4 in Figure 6-8c). When the shear plane encounters a triple junction and is stopped by a grain that is not in its alignment, the shear band follows its path by intragranular slip in the prolongation of the shear plane. For example, a stacking fault left behind a partial dislocation can be seen in grain 2 in the prolongation of the shear plane in Figure 6-8b. Subsequently, the newly created stacking faults are found to nucleate mechanical twins, which grow under the applied shear stress. Mechanical twinning has also been observed in nanocrystalline Al under indentation by Chen et al. [28] (This result therefore suggests that there is good agreement between simulation and experimental data).

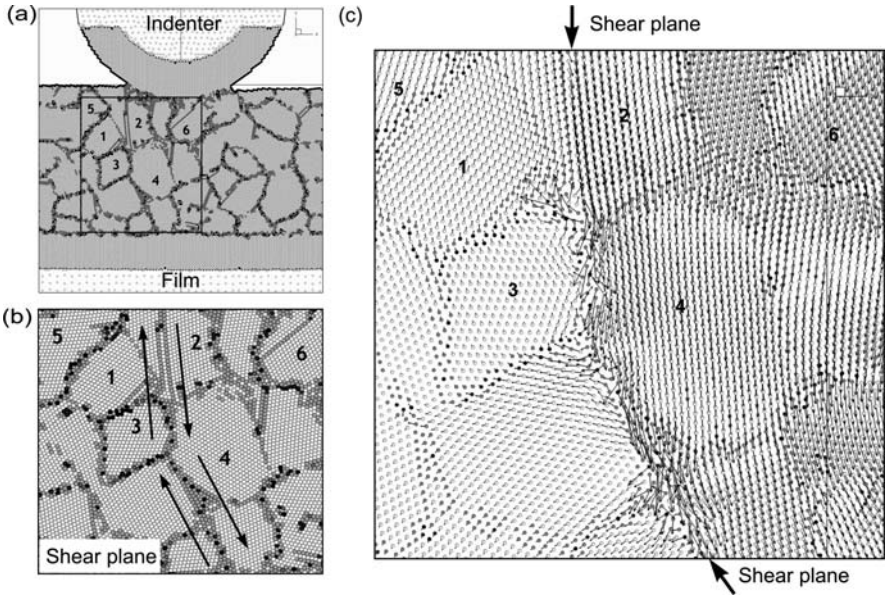


Figure 6-8. Formation of a thin shear band in a 5-nm-grain-size nanocrystalline Al during indentation by QC method. (**a**) Partial view of the contact interface and location of the grain cluster associated with the shear band. (**b**) Close up view of the shear plane. (**c**) Magnitude and direction of atomic displacements between two loading increments shown by arrows. Reprinted from [47] with permission from Elsevier

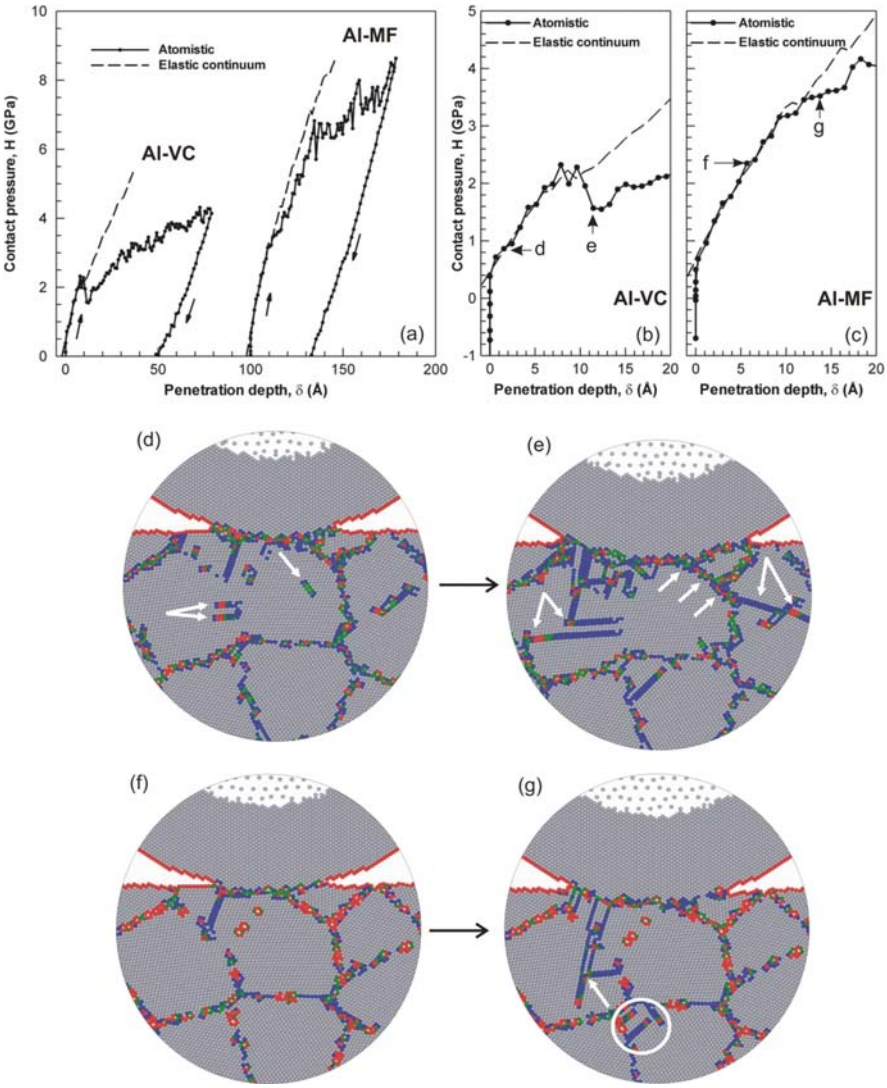


Figure 6-9. Effects of EAM potential on the mean contact pressure as a function of penetration depth in a 7 nm-grain-size QC simulation. (**a**)–(**c**) Loading and unloading responses with Voter-Chen (Al-VC) and Mishin-Farkas (Al-MF) EAM potentials for Al. Snapshots showing the deformation in the film near the contact region for different stages of indentation for (**d**)–(**e**) the Al-VC potential, and (**f**)–(**g**) the Al-MF potential. Adapted from [44] with permission from Elsevier

Furthermore, Figure 6-9 represents the evolution of the mean contact pressure as a function of penetration depth for both the Al-VC and Al-MF potentials. A fitting between atomistic simulation and elastic continuum theory [44] shows an excellent agreement in the initial portion of the curves. Both curves also show clear

evidence of flow serration at large depths of indentation, which is consistent with the formation of shear bands. In Figure 6-9, however, the plastic flow serration is somewhat less intense with the Al-VC potential than the Al-MF potential. In the latter, a clear change in slope occurs for a penetration depth of 34 Å, which corresponds to the propagation of an extensive shear band by intragranular slip through several grains (not shown). Furthermore, Figures 6-9b and c reveal that the contact pressures are almost identical between atomistic results and elastic theory until the two curves reach a depth of 10–15 Å. The divergence between elastic and atomistic curves marks the onset of plasticity from a macroscopic, continuum standpoint. These curves also provide direct evidence that the onset of plasticity occurs at significantly larger contact pressure in the Al-MF potential than the Al-VC potential. The pressure difference becomes even more significant at the end of the indentation process, where contact pressures of 4.3 and 8.6 GPa are observed for the two potentials, respectively. Investigation of the depth of the residual impression after unloading also shows some significant differences in the constitutive response between the two potentials.

The change in nanoindentation behavior can be largely interpreted from fundamental differences in plastic deformation mechanisms at GBs as a function of potential. While the nucleation of a few dislocations takes place at the tip/substrate contact interface during the elastic portion of the contact curves with both potentials (Figures 6-9d and f), more GB-mediated activity in the substrate occurs with the Al-VC potential. Here, GB-mediated plasticity is characterized by one of the following mechanisms: (1) the emission of partial dislocations and twins emanating from GBs, along with their propagation through intragranular slip, (2) GB sliding and grain rotation, and (3) stress-driven GB migration coupled to shear deformation.

6.5. MECHANISMS OF GRAIN BOUNDARY MOTION DURING CONTACT PLASTICITY

Several mechanisms for grain growth have been investigated in nanocrystalline metals. These can primarily be divided into two categories: Thermally-driven mechanisms and stress-assisted mechanisms (Figure 6-10). The first category is related to curvature-driven GB motion [66] (Figure 6-10a). During this process, the GB tends to move towards the center of the curvature to reduce the total area of GBs, and thus the energy of the system [67]. Another thermally-driven mechanism is GB atom diffusion [68], during which the atoms jump in the crystal into point vacancies, creating a new vacancy in the process.

The last two mechanisms in Figure 6-10 represent stress-assisted mechanisms, which can make grain growth at cryogenic temperatures possible [31]. The first mechanism corresponds to rotation-induced grain coalescence [67], during which one grain rotates in order to match the orientation of a neighbor grain, thus forming a single larger grain (Figure 6-10d). This process is often associated with GB sliding [69]. The second mechanism of strain-driven grain growth is related to shear-coupled GB motion (Figure 6-10e). In this case, the normal motion of

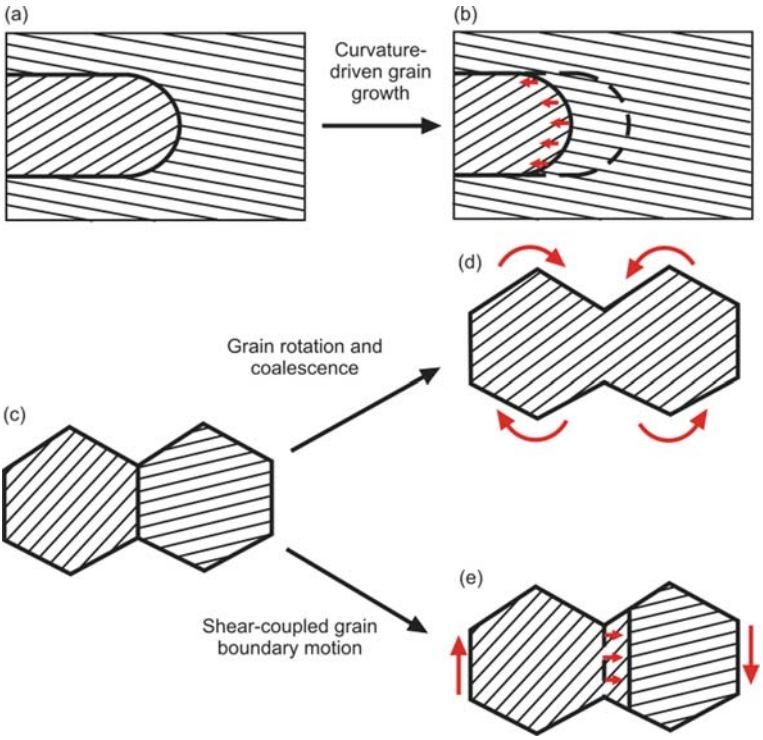


Figure 6-10. Schematic illustrations of different mechanisms of GB motion in metals. (**a**)–(**b**) Thermally-driven GB motion. (**c**) Strain-driven GB motion by either (**d**) grain rotation and coalescence or (**e**) shear-coupled GB migration

GBs results from a shear strain applied tangentially to them and causing tangential motion, or coupled motion [69–72].

To illustrate such grain growth mechanisms during indentation, Figure 6-11 presents the indentation of a 7-nm-grain-size Al film by a spherical tip from MD simulation at 300 K. In this figure, the GB atoms and lattice defects appear in white color, and the other atoms are colored according to the grain they belonged to at the beginning of the simulation. Several major grain growth events are visible in Figure 6-11: between grains 1 and 2, between grains 3 and 4, for grains 5, 6, 7 and 8, which all appear to be strain-driven as demonstrated below.

The grain growth events occurring between grains 1 and 2 and between grains 3 and 4, starts very early in the indentation process. A closer inspection of the evolution of these grains shows that the misorientation angles at the GB are very small before relaxation, which makes them low-angle GBs. After relaxation, the grains slightly rotate, until the misorientation angle is low enough to have the atoms at the interface in perfect FCC arrangement. The corresponding mechanism of GB migration is therefore rotation-induced grain coalescence.

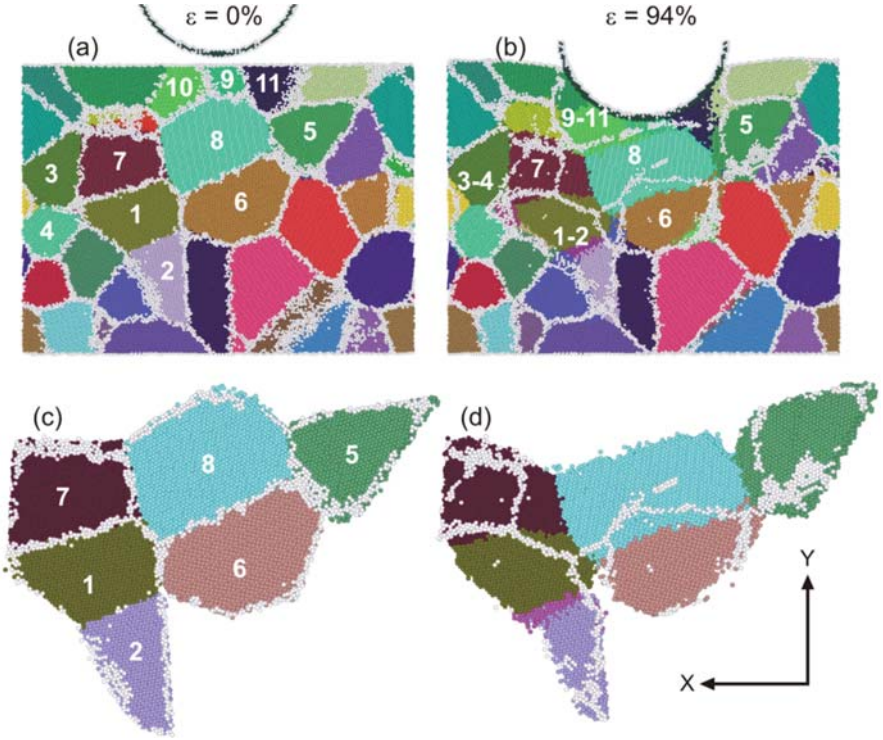


Figure 6-11. Three-dimensional MD simulation of microstructure evolution in nanocrystalline Al indented by a 9 nm-radius Al tip. General view obtained (**a**) after relaxation and before indentation ($\varepsilon = 0\%$) and (**b**) at maximum indentation ($\varepsilon = 94\%$). (**c**) Close-up view of grains 1, 2, 5, 6, 7 and 8 when (**c**) $\varepsilon = 0\%$ and (**d**) $\varepsilon = 94\%$ [48]

The deformation process in grains 5, 6, 7 and 8 is clearly different from that in grains 1, 2, 3 and 4, because the boundaries are followed moving across neighbor grains (Figure 6-11d). We find that the rate of GB migration can be strongly decreased as the tip radius decreases which tends to indicate that this mechanism depends on the local strain. Furthermore, the shape of grain 7 evolves from originally square to trapezoidal after migration of the interface between grains 7 and 8. These results suggest that the process of GB migration is coupled to shear deformation in this case. In this process, grains 6 and 8 have grown, while grains 5 and 7 have notably decreased in diameter. Also a new grain was grown at the triple junction of grains 1, 6 and 8.

However, MD simulation makes the analysis of strain-driven GB motion difficult due to the implications of thermal effects. In contrast, QC simulations, which operate at zero temperature, may readily eliminate those effects. For example, Figure 6-12 shows the motion of the interface separating grains 3 and 5, which corresponds to a high-angle GB ($\psi + \psi' = 160.7°$), from the QC simulation presented in Figure 6-5 with the use of the Al-VC potential.

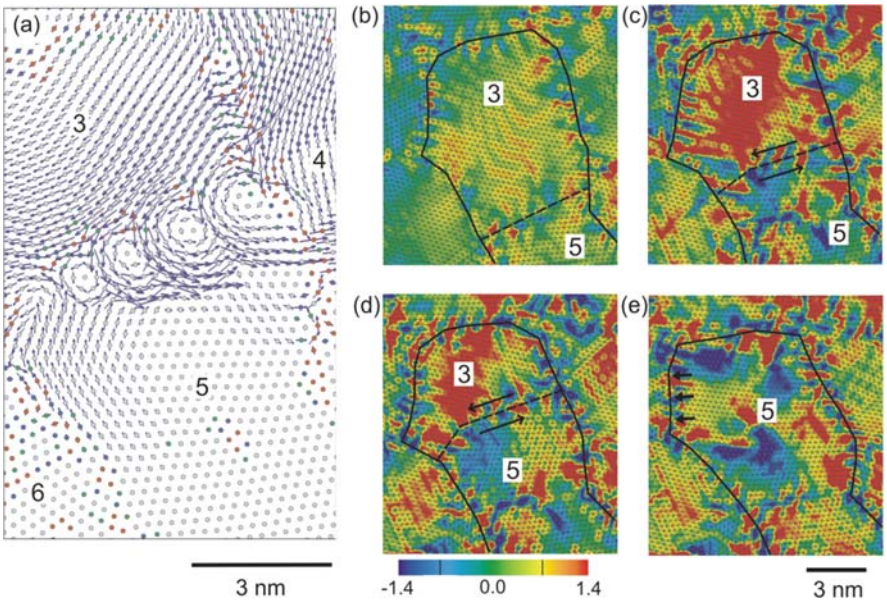


Figure 6-12. QC simulation of shear-coupled GB motion in nanocrystalline Al during indentation. (**a**) Local lattice rotations at GB forming vortex-like patterns. (**b**)–(**e**) Snapshots showing the evolution of the atomic-level shear stress relative to the orientation of the interface between grains (in units of GPa). The boundaries of grains 1 and 2 have been highlighted for clarity. Adapted with permission from [46].

Each stress map represented in Figure 6-12 corresponds to a different loading level taken from before the start of the migration to the end of the GB motion. It is found that the interface migration process causes a strong stress relief in the GB network. Furthermore, the sign of the stress, indicated by the color blue or red in Figure 6-12 indicates that the GB between grains 3 and 5 is under shear stress, one side being under positive stress, and the other under negative stress. In this GB migration process, it is clear that the size of grain 5 has increased at the expense of grain 3. Therefore, this analysis shows clear evidence that the mechanism of stress-assisted grain growth corresponds to shear-coupled motion.

Furthermore, Figure 6-13 shows that GB migration is only observed for the Al-VC simulation. The original position of the GB before indentation is indicated by a dashed line for reference. The GB in the Al-VC simulation has moved about 34 Å into grain 3, whereas no significant differences are found with the Al-MF simulation, other than some minor rearrangements of GB atoms.

In summary, the mechanism of stress-driven grain growth was found associated with shear-coupled GB motion, rather than by rotation-induced grain coalescence. We can therefore conclude that the plastic flow is found enhanced in the case of the Al-VC simulation because of the increased GB deformation activity in the form of GB sliding and coupled GB motion at both 0 and 300 K. This finding could therefore

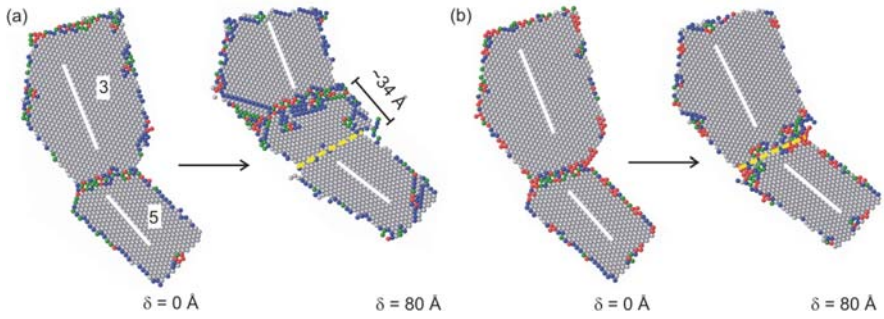


Figure 6-13. Effects of EAM potential on the shear-coupled GB migration between grains 3 and 5 as shown in Figure 6-6, as a function of penetration depth δ. (**a**) Voter-Chen potential. (**b**) Mishin-Farkas potential. Reprinted from [44] with permission from Elsevier

suggest a new means to control stress-assisted grain growth mechanisms by altering the local structure and energies at GBs.

6.6. CONCLUDING REMARKS

In this chapter, we describe two computational approaches using either MD or QC methods to investigate the complexity of plastic deformation and yielding in nanocrystalline GB networks under nanoscale contact. These approaches show that the interatomic potentials play a key role in the prediction of equilibrium structures and energies at GBs, which in turn has a strong influence on the mechanisms of plasticity mediated by GBs, such as GB-mediated crystal slip, GB sliding, grain rotation and GB migration. It is also demonstrated that shear-coupled GB motion due to the strain applied by penetrating tips, is an important mechanism of grain growth and plasticity at both absolute zero and room temperature.

An outlook for future research in this area can be summarized by asking two fundamental questions. First, can the addition of impurities significantly influence the simulation of GB-mediated plasticity in nanocrystalline metals? It is acknowledged that solute impurities, like H and O impurities in Al, have strong impact on stacking fault and GB energies, and plasticity in FCC metals [32, 73–75]. To this end, proper numerical formalisms, such as local chemical potential [75], should be developed to include impurities in multiscale simulations. Second, how do the geometry and deformation of tips influence plasticity induced by nanoscale contact in nanocrystalline materials? Clearly, present simulations do not take into account the finite-temperature deformation of the tips. Atomistic simulation on this aspect may provide more fundamental understanding of tip-film interactions, which may ultimately help engineers explore new routes for high-throughput nanomanufacturing technology, such as nanoimprint lithography [76].

ACKNOWLEDGMENT

Support from National Science Foundation CAREER program (grant no DMR-0747658) and the computational resources provided by the Vermont Advanced Computing Center, which is supported by NASA (grant no NNX 06AC88G), are gratefully acknowledged.

REFERENCES

1. Basrour, S. and Robert, L. (2000) Mater. Sci. Eng. A 288: 270–274.
2. Gobet, J., Cardot, F., Bergqvist, J. and Rudolf, F. (1993) J. Micromech. Microeng. 3: 123–130.
3. Larsen, K.P., *et al*. (2003) Sens. Actuators A 103: 156–164.
4. Martinez, S., *et al*. (2002) Sens. Actuators A 99: 41–44.
5. Pasa, A.A. and Schwarzacher, W. (1999) Phys. Stat. Sol. A 173: 73.
6. Krauss, A.R., *et al*. (2001) Diam. Relat. Mater. 10: 1952–1961.
7. Tao, S. and Li, D.Y. (2006) Nanotechnology 17: 65–78.
8. Schiotz, J. and Jacobsen, K.W. (2003) Science 301: 1357–1359.
9. Trelewicz, J.R. and Schuh, C.A. (2007) Acta Materialia 55: 5948–5958.
10. Chang, S.Y. and Chang, T.K. (2007) J. Appl. Phys. 101: 033507.
11. Tsuchiya, T., Tabata, O., Sakata, J. and Taga, Y. (1996) IEEJ Trans. 116: 441.
12. Sharpe, W.N., Yuan, B. and Edwards, R.L. (1997) J. Microelectromech. Syst. 6: 193–199.
13. Chasiotis, I. and Knauss, W.G. (2002) Exp. Mech. 42: 51–57.
14. Hemker, K.J. and Sharpe, W.N. (2007) Annu. Rev. Mater. Res. 37: 93–126.
15. Espinosa, H.D., Prorok, B.C. and Fischer, M. (2003) J. Mech. Phys. Solids 51: 47–67.
16. Nieman, G.W., Weertman, J.R. and Siegel, R.W. (1989) Scripta Metall. 23: 2013–2018.
17. Elsherik, A.M., Erb, U., Palumbo, G. and Aust, K.T. (1992) Scripta Metall. Mater. 27: 1185–1188.
18. Fougere, G.E., Weertman, J.R. and Siegel, R.W. (1995) Nanostruct. Mater. 5: 127–134.
19. Qin, X.Y., Wu, X.J. and Zhang, L.D. (1995) Nanostruct. Mater. 5: 101–110.
20. Farhat, Z.N., Ding, Y., Northwood, D.O. and Alpas, A.T. (1996) Mater. Sci. Eng. A 206: 302–313.
21. Malow, T.R., Koch, C.C., Miraglia, P.Q. and Murty, K.L. (1998) Mater. Sci. Eng. A 252: 36–43.
22. Sanders, P.G., Eastman, J.A. and Weertman, J.R. (1997) Acta Mater. 45: 4019–4025.
23. Gang, T. and Sansoz, F. (2007) Unpublished Research.
24. Minor, A.M., *et al*. (2006) Nat. Mater. 5: 697–702.
25. Yang, B. and Vehoff, H. (2007) Acta Materialia 55: 849–856.
26. Andrievski, R.A., Kalinnikov, G.V., Jauberteau, J. and Bates, J. (2000) J. Mater. Sci. 35: 2799–2806.
27. Van Vliet, K.J., Tsikata, S. and Suresh, S. (2003) Appl. Phys. Lett. 83: 1441–1443.
28. Chen, M.W., *et al*. (2003) Science 300: 1275–1277.
29. Jin, M., Minor, A.M., Stach, E.A. and Morris, J.W. (2004) Acta Materialia 52: 5381–5387.
30. Zhang, K., Weertman, J.R. and Eastman, J.A. (2004) Appl. Phys. Lett. 85: 5197–5199.
31. Zhang, K., Weertman, J.R. and Eastman, J.A. (2005) Appl. Phys. Lett. 87: 061921.
32. Gianola, D.S., *et al*. (2006) Acta Materialia 54: 2253–2263.
33. Hall, E.O. (1951) Proc. Phys. Soc. B 64: 747–753.
34. Petch, N.J. (1953) J. Iron Steel Inst. 174: 25–28.
35. Schiotz, J., Di Tolla, F.D. and Jacobsen, K.W. (1998) Nature 391: 561–563.
36. Feichtinger, D., Derlet, P.M. and Van Swygenhoven, H. (2003) Phys. Rev. B 67: 024113.
37. Ma, X.L. and Yang, W. (2003) Nanotechnology 14: 1208–1215.
38. Lilleodden, E.T., Zimmerman, J.A., Foiles, S.M. and Nix, W.D. (2003) J. Mech. Phys. Solids 51: 901–920.
39. Hasnaoui, A., Derlet, P.M. and Van Swygenhoven, H. (2004) Acta Materialia 52: 2251–2258.

40. Jang, H. and Farkas, D. (2004) Mater. Res. Soc. Symp. Proc. 821: P8.17.1–P8.17.6.
41. Saraev, D. and Miller, R.E. (2005) Model. Simulat. Mater. Sci. Eng. 13: 1089–1099.
42. Kim, K.J., Yoon, J.H., Cho, M.H. and Jang, H. (2006) Mater. Lett. 60: 3367–3372.
43. Szlufarska, I., Nakano, A. and Vashishta, P. (2005) Science 309: 911–914.
44. Dupont, V. and Sansoz, F. (2008) Acta Mater. 56: 6013–6026.
45. Luan, B.Q., *et al*. (2006) Phys. Rev. E 74: 11.
46. Sansoz, F. and Dupont, V. (2006) Appl. Phys. Lett. 89: 111901.
47. Sansoz, F. and Dupont, V. (2007) Mater. Sci. Eng. C 27: 1509–1513.
48. Dupont, V. (2008) Ph.D. Thesis. University of Vermont.
49. Zimmerman, J.A., Gao, H.J. and Abraham, F.F. (2000) Model. Simulat. Mater. Sci. Eng. 8: 103–115.
50. Plimpton, S. (1995) J. Comput. Phys. 117: 1–19.
51. Miller, R.E. and Tadmor, E.B. (2002) J. Comput. Aid Mater. Des. 9: 203–239.
52. Rapaport, D.C. (2004) The Art of Molecular Dynamics Simulation. Cambridge University Press, Cambridge, UK, 549.
53. Hoover, W.G. (1985) Phys. Rev. A 31: 1695–1697.
54. Shenoy, V.B., *et al*. (1999) J. Mech. Phys. Solids 47: 611–642.
55. Sansoz, F. and Molinari, J.F. (2007) Thin Solid Films 515/6: 3158–3163.
56. Johnson, K.L. (1985) Contact Mechanics. Cambridge University Press, Cambridge, UK, 452.
57. Voronoi, G.F. (1908) J. Reine Undangew. Math. 134: 199–287.
58. Kelchner, C.L., Plimpton, S.J. and Hamilton, J.C. (1998) Phys. Rev. B 58: 11085–11088.
59. Ackland, G.J. and Jones, A.P. (2006) Phys. Rev. B 73: 054104.
60. Daw, M.S. and Baskes, M.I. (1983) Phys. Rev. Lett. 50: 1285–1288.
61. Mishin, Y., Farkas, D., Mehl, M.J. and Papaconstantopoulos, D.A. (1999) Phys. Rev. B 59: 3393–3407.
62. Voter, A.F. and Chen, S.P. (1987) Mater. Res. Soc. Symp. Proc. 82: 175.
63. Sansoz, F. and Molinari, J.F. (2005) Acta Materialia 53: 1931–1944.
64. Sansoz, F. and Molinari, J.F. (2004) Scripta Materialia 50: 1283–1288.
65. Van Swygenhoven, H., Derlet, P.M. and Froseth, A.G. (2004) Nat. Mater. 3: 399–403.
66. Zhang, H., Upmanyu, N. and Srolovitz, D.J. (2005) Acta Materialia 53: 79–86.
67. Haslam, A.J., *et al*. (2001) Mater. Sci. Eng. A 318: 293–312.
68. Huntington, H.B. and Seitz, F. (1942) Phys. Rev. 61: 315.
69. Cahn, J.W., Mishin, Y. and Suzuki, A. (2006) Acta Materialia 54: 4953–4975.
70. Cahn, J.W. and Taylor, J.E. (2004) Acta Materialia 52: 4887–4898.
71. Gutkin, M.Y., Mikaelyan, K.N. and Ovid'ko, I.A. (2008) Scripta Materialia 58: 850–853.
72. Gutkin, M.Y. and Ovid'ko, I.A. (2005) Appl. Phys. Lett. 87: 251916.
73. Lu, G., Zhang, Q., Kioussis, N. and Kaxiras, E. (2001) Phys. Rev. Lett. 8709: 095501.
74. Lu, G., *et al*. (2002) Phys. Rev. B 65: 064102.
75. Elsener, A., Politano, O., Derlet, P.M. and Van Swygenhoven, H. (2008) Model. Simulat. Mater. Sci. Eng. 16: 025006.
76. Chou, S.Y., Krauss, P.R. and Renstrom, P.J. (1996) J. Vac. Sci. Technol. B 14: 4129–4133.

CHAPTER 7

SILICON NANOWIRES: FROM EMPIRICAL TO FIRST PRINCIPLES MODELING

RICARDO W. NUNES[1] AND JOÃO F. JUSTO[2]

[1]*Departamento de Física, ICEX, Universidade Federal de Minas Gerais, CP 702, 30123-970, Belo Horizonte, MG, Brazil,e-mail: rwnunes@fisica.ufmg.br*

[2]*Escola Politécnica, Universidade de São Paulo, CP 61548, CEP 05424-970, São Paulo, SP, Brazil, e-mail: jjusto@Ime.usp.br*

Abstract: Silicon nanowires have been the subject of intense investigation over the last decade. The experimental realization of nanowire configurations with a wide range of diameters, lengths, and surface types leads us to envision a wealth of applications, running from selective sensors of rapid response to electronic devices. In this period, theoretical modeling has helped to understand the electronic, mechanical, optical and transport properties of nanowires and to explore applications of such properties in the context of the current electronic technologies. These modern theoretical calculations have reached a point where realistic description of materials properties are provided by computational simulations. Carefully constructed empirical potentials provide a good description of silicon-nanowire energetics, making possible investigations of the stability of silicon wires with different surface terminations (or facets) for a given family of nanowires. Simulations using empirical potentials have also been employed to examine the thermal and mechanical stability of silicon nanowires, and also the response under external load. In the latter case, there is and indication that the response under load of silicon nanowires is different from the bulk, to the extent that crack propagation is suppressed and healing at the crack is mediated by surface effects. Investigation of nanowire properties are also within the reach of tight-binding and first principles methodologies, that have been used to examine the effects of quantum confinement on the nanowire electronic, transport and structural properties, as well as the nature of their surface states. These methods have also been used to examine the possibility of structural transitions of very thin silicon wires, induced by surface effects

Keywords: Nanowires, Atomistic modeling

7.1. INTRODUCTION

Nanoscience and nanotechnology have seen a remarkable development over the last decade, mostly because substantial advances in experimental techniques have allowed to manipulate matter at the nanometer scale. For example, the current

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 173–191.
DOI 10.1007/978-1-4020-9785-0_7,

challenge in developing faster and more efficient devices has driven researchers to explore the limits of materials properties, unavoidably leading to the nanoworld [1]. In this context, one of the main goals is to understand and manipulate one-dimensional (1D) nanostructured systems, which are envisaged as fundamental building blocks for constructing nanoscale electronic devices [2–6]. A variety of prototype devices have already been constructed. To name only a few, under the risk of ignoring some of the voluminous literature that has accumulated on this topic, reported prototype-device fabrication include: electron memory cells [7, 8], passive diode structures using crossed p- and n-type silicon nanowires and active bipolar transistors based on heavily doped nanowires [9], logic gates built with integrated nanoscale field-effect transistors [10], three terminal devices [11], photonic devices [3], nanowire based lasers [12, 13], and chemical and biological sensors [14–16]. Besides the interest for device applications, studies on nanowires offer the possibility of understanding, fundamentally, the roles of dimensionality and size in optical, electrical, magnetic, mechanical and transport properties of low-dimensional condensed-matter systems.

Along with carbon nanotubes [17], semiconducting nanowires are among the most extensively studied nanostructured systems. To a large extent, silicon has remained the material of choice for the electronic device technology, and silicon nanowires (SiNWs), in particular, have attracted great interest, since it is expected that any proposed silicon-based nano-devices should be "easily" integrated into the existing silicon-based electronic technology. When properly functionalized, SiNWs may work as electronic devices, such as diodes or transistors [18]. Moreover, narrow SiNWs present quantum confinement, which should allow gap engineering and the possibility of incorporating silicon based optical devices in integrated circuits [19–21]. Full silicon integration into nanotechnology requires understanding the structural and electronic properties of SiNWs and devising controlled processes to build SiNWs with tailored properties.

Starting in the mid 1990's, several routes have been demonstrated for the synthesis of semiconducting nanowires. Perhaps the first reported work on silicon nanowires synthesis, as such, is that of Liu and collaborators, which employed a combination of electron beam lithography, reactive ion etching, and dry thermal oxidation [22, 23]. Further studies employed a variety of synthesis techniques, such as etching [24], nanolitography [25], scanning tunnelling microscopy (STM) manipulation of silicon atoms on substrates [26], deposition of nanowire material into porous templates [27, 28], laser ablation [29, 30], and deposition or sublimation of silicon vapor [29, 31–35], among others. In order to devise SiNWs with pre-determined properties, it is important to develop procedures that enable controlling the growth direction and surface types. The search for better control of nanowire surface morphology and crystalline orientation has led to synthesis protocols based on either vapor-liquid-solid or super-critical-solution liquid-solid processes, in which nanowires grow usually from a metallic nanoparticle catalyst [36–46]. Several authors have succeeded in getting nanowires with a wide range of diameters and along several crystallographic lattice directions: Holmes et al. have grown SiNWs along the ⟨100⟩ and ⟨110⟩ lattice directions [37]. Ma et al. reported

ultra-thin hydrogen-terminated wires along ⟨100⟩ directions [46]. Wu et al. reported SiNWs along ⟨110⟩, ⟨111⟩, and ⟨112⟩ lattice directions [47].

With these experimental developments leading to the synthesis of SiNWs from a variety of methods, it was only natural that the expertise gained by theorists in addressing the structural and electronic properties of silicon clusters [48, 49], using ab initio, tight-binding (TB) and classical potential methodologies, would be applied to the study of these properties in SiNWs. Indeed, over the last few years, theoretical investigations on the properties of SiNWs have been performed using those methodologies or a combination of them. Among the pioneers, Menon and Richter [50] employed TB molecular dynamics to investigate the structural stability of Si wires, concluding that, at variance with small- and intermediate-sized silicon clusters, which undergo extensive reconstruction and relaxation down to the core atoms, a fourfold-coordinated crystalline nanowire core is stable when surrounded by a reconstructed surface of threefold coordinated atoms, drawn from the most stable bulk reconstructions. Single-walled nanotubes (SWSiNT) were also considered as possible quasi-one-dimensional forms of silicon. Several works examined the structural and electronic properties of SWSiNTs [51–56]. However, despite some works suggesting otherwise [54], more likely, in the case of silicon, the overlap of π-electrons is not strong enough to stabilize the planar graphene form, and a carbon-like SWSiNT should be mechanically and thermally unstable [52, 55, 57], and even at zero temperature puckering should be considerable and a substantial amount of sp^3-hybridized orbitals would mix into the sp^2 network [52, 55, 56]. Ab initio calculations by Dumitrica et al. [58] indicated that hollow nanotube-like structures can be stabilized by endohedral metal atoms. Other types of hollow SiNT structures have been considered, with facetted cross sections [59], variable wall thickness [60, 61], or with hydrogen-passivated surfaces [62]. These may be considered to fall in the category predicted by Menon and Ritcher, to the extent that they represent sp^3-hybridized structures with either tetrahedral or hexagonal symmetry, and with either reconstructed or saturated surfaces.

In the last few years, theoretical research has focused on sp^3 silicon-bulk derived structures [63–81], given the experimental observation of the prevalence of nanowires oriented along the ⟨110⟩, ⟨112⟩, and ⟨111⟩ bulk crystalline directions [47]. A number of different possibilities, with prevailing sp^3 structures, have been considered: trigonal-prism geometries [64]; fivefold and sixfold cross-section symmetries both crystalline [77] and polycrystalline [63, 65]; and cage-like structures [31, 67–69, 71, 75]. The energetics and orientational dependence of electronic, mechanic, and surface properties in the bulk-derived geometries have been addressed by a number of authors [66, 70, 72–74, 76, 79, 81]. The works of Ng et al. on gap reduction upon surface functionalization [74]; of Lu et al. on hydrogen-chemical-potential dependence of the stability of passivated wires [76]; of Rurali et al. on charge injection effects upon structural features [79]; and of Wu et al. on tapering and charge inhomogeneities effects [81], point in the direction of important issues related to the future SiNW integration. Another issue that deserves a full address is the stability both mechanical and thermal of the various alternative nanowire structures proposed in these works. For example, identification of possible

soft-mode phonons, and of the strain-state dependence of these modes, is required in order to address the issue of mechanical stability of these geometries against structural transformation.

In the following, we outline some recent theoretical investigations that employed atomistic simulations to address the energetics and structural properties of SiNWs.

7.2. METHODOLOGICAL CONSIDERATIONS

Theoretical investigations, based on atomistic simulations, have become a powerful tool to understand and/or predict materials properties. This field has blossomed as a result of intensive research and the development of precise methodologies over the last decades, which now enable researchers to describe materials properties with an unprecedented precision. The central element in the investigation of materials properties is the level in which interatomic interactions are described, ranging from pure first principles models, based on quantum mechanics with no fitting parameters, to semi-empirical ones, in which quantum mechanics is still considered but with a few fitting parameters, to empirical models, based on classical mechanics using functional forms fitted to experimental results.

First principles (or ab initio) methods take into account all the particles (electrons and nuclei) of the system and their fundamental Coulomb interactions [82–84]. In the context of condensed matter physics and materials science, first principles approaches are usually based on the density functional theory (DFT) [82, 83]. In a few words, in DFT it is shown that the ground-state energy of a many-body system of electrons and nuclei is a functional of the electronic density of the system, and an operational scheme is proposed that maps the many-body problem into an effective single-particle set of equations, known as Kohn-Sham equations. The mapping is exact, and the only fundamental approximation needed in practical implementations of this scheme is due to the unknown form of the so-called exchange and correlation functional, which contains the contribution of quantum-fluctuations to the mutual electronic Coulomb interactions, and also a many-body correction to the kinetic energy. This has proved a very successful approach to address the ground-state properties of a large variety of condensed-matter systems [84].

While first principles approaches provide a precise description of the interatomic interactions and consequently of the materials properties, the computational costs involved scale with some power law on the number of atoms in the system. As a result, simulations are restricted to systems involving only up to, at most, a thousand atoms, which is not enough to capture many important atomistic and mesoscopic phenomena. Approximations, leading to less expensive models, that make possible to treat systems involving a higher number of atoms have been implemented. Semi-empirical models are still computationally intensive but allow investigations of systems with thousands of atoms. These are based on a local minimal-basis description of the quantum mechanical interactions, in the so-called tight-binding (TB) approximation. The strategy, in this case, is to constrain the computational scale of the problem by restricting the electronic matrix elements, that constitute

the Hamiltonian of the electronic interactions, to near-neighbor interactions and to a sub-space of the Hilbert space that spans the valence bands (or highest-occupied molecular orbitals in a molecular system) and a few of the lower conduction bands (or lowest-unoccupied molecular orbitals). Moreover, the multi-center integrals needed to evaluate the Hamiltonian matrix elements are not computed. Rather, the matrix elements are parameterized such that results for a set of selected properties, in a reference configuration, agree with either experimental or first principles results.

A more radical approach is to describe the interactions by classical potentials, in which the electronic effects are taken into account only implicitly, allowing to treat systems with millions of atoms. The gain in computational efficiency with semi-empirical and empirical methods comes with a price: a considerably poorer description of the microscopic phenomena, such that those models should be used with caution, staying within their range of validity. A very useful approach to describe nanosystems has been to combine methods with different levels of approximation, in the so-called multi-scale methods that combine empirical, semi-empirical and first principles methods [85, 86].

In the following section, we describe the main theoretical ingredients of empirical and semi-empirical models, that have been employed to investigate SiNW properties. A review on ab initio methodologies can be found elsewhere [84].

7.2.1. Empirical Models

The cohesive energy, $E_c(\mathbf{R}_n,\mathbf{r}_m)$, quantifies the cohesive strength of a material, and is given in terms of the degrees of freedom of the N_a nuclei ($[\mathbf{R}_n]$) and N_e electrons ($[\mathbf{r}_m]$) of the system. While it could be computed by solving the quantum mechanical equations for the electrons of the system, approximations can be used to describe cohesion with less expensive methods. One strategy is to employ an approximate description of the electronic states, but still keeping the electronic degrees of freedom explicitly. One of these approaches, the TB method, provides a realistic description of interatomic bonding, although it is still computationally intensive. This method will be described in Section 7.2.2. An extreme approach is to remove all the electronic degrees of freedom, and E_c would be given by $E_c(\mathbf{R}_n,\mathbf{r}_m) \approx E_c(\mathbf{R}_n)$. In this case, the electronic effects would be considered only implicitly, in the construction of the functional form of the empirical potentials.

For materials in which the covalent interactions prevail, several empirical potentials have been developed over the years. Only for silicon, there are more than forty models in the literature, many of them have been extensively used and tested [87, 88]. Three empirical potentials have been intensively used to model silicon properties: the Stillinger-Weber [89], the Tersoff [90], and the EDIP [91] models. Specifically for SiNWs, all these potentials have been used to study the structural properties and stability, and have been compared recently [92]. Overall, all three models provide a reasonably realistic description of the structural properties of these wires.

Cohesive energy can be written as a function of the atomic arrangement, in terms of a many-body expansion [93]

$$E_{\mathrm{c}} = \sum_{i}^{N_{\mathrm{a}}} V_1(\mathbf{R}_{\mathrm{i}}) + \sum_{i,j}^{N_{\mathrm{a}}} V_2(\mathbf{R}_{\mathrm{i}},\mathbf{R}_{\mathrm{j}}) + \sum_{i,j,k}^{N_{\mathrm{a}}} V_3(\mathbf{R}_{\mathrm{i}},\mathbf{R}_{\mathrm{j}},\mathbf{R}_{\mathrm{k}}) + \cdots , \qquad (7\text{-}1)$$

in which the sums are over all the N_{a} atoms of the system. In principle, E_{c} could be determined by an infinite many-body expansion, but the computational cost scales with $N_{\mathrm{a}}^{\mathrm{l}}$, where l the order in which the expansion is truncated. The one-body terms (V_1) are generally neglected, but the two-body (V_2) and three-body (V_3) terms carry most of the relevant effects underlying bonding. While the V_2 and V_3 have a simple physical interpretation, intuition for higher order terms is not so straightforward, and most models have avoided such terms. Truncation of this expansion up to only two-body terms generally fails in capturing the essential properties of covalent systems; higher expansion terms, i.e. the many-body interactions, are necessary. The many-body effects [93] could be introduced in E_{c} by several procedures: inside the two-body expansion (pair functionals), by an explicit many-body expansion (cluster potentials), or a combination of both (cluster functionals). Models which have been successfully developed to describe covalent systems fit into one of these categories. All three models mentioned earlier [89–91] include these many-body effects in different approaches. The Stillinger-Weber [89] model can be classified as a cluster potential, with an explicit three-body expansion, while the Tersoff [90] and EDIP [91] models can be classified as cluster functionals.

7.2.2. Semi-Empirical Models

In semi-empirical TB methods, the quantum-mechanical nature of the system is retained. The terminology "semi-empirical" is related to the traditional way of setting up TB Hamiltonians, by fitting the Hamiltonian matrix elements of a minimal-set basis orbitals to experimentally measured band-structures, in a reference configuration. The minimal-set basis orbitals are chosen such as to span the sub-space of the valence band and a few of the lowest conduction bands. More general formulations which included efficient scaling laws have been introduced [94], and total energies have also been incorporated, with the addition of a repulsive classical potential term, which accounts for the Coulomb and overlap repulsive ion-ion (nuclei + core electrons) interactions. These tight-binding total energy formulations (TBTE) also include matrix elements scaling laws, allowing for structural relaxation and molecular dynamics simulations of systems with several thousand atoms, with very good accuracy.

In a TBTE scheme, the total energy of a system containing N_{e} electrons, as a function of the $\{\mathbf{R}_{\mathrm{n}}\}$ nuclear positions, is given by

$$E_{\mathrm{tot}}\left[\{\mathbf{R}_n\}\right] = \sum_{i}^{N_{\mathrm{occ}}} \varepsilon_i + E_{\mathrm{rep}}\left[\{\mathbf{R}_n\}\right] \ . \qquad (7\text{-}2)$$

In this expression, E_{tot} is the total energy, the sum is over the N_{occ} eigenvalues (ε_i) of the occupied electronic states, and E_{rep} is the repulsive term. The occupied

states and eigenvalues are obtained by diagonalization of a TB Hamiltonian, whose matrix elements, in their more usual form, are written as:

$$H^{n,m}_{\alpha,\beta}\left(\mathbf{R}_\mathrm{n}-\mathbf{R}_\mathrm{m}\right)=H^{0}_{\alpha,\beta}\times f_\mathrm{sc}\left(\mathbf{R}_\mathrm{n}-\mathbf{R}_\mathrm{m}\right)\ , \tag{7-3}$$

where $H^{n,m}_{\alpha,\beta}$ is the matrix element between local-basis function ϕ_α centered in the n-th nuclei and basis function ϕ_β centered in the m-th nuclei, $H^{0}_{\alpha,\beta}$ is the value of this matrix element in some reference configuration, and f_sc is the scaling law that gives the dependence of $H^{n,m}_{\alpha,\beta}$ in the interatomic distance. The reference values of these interactions and the scaling functions are determined by fitting the TB results to a chosen set of reference values, either from experiments or from first principles calculations.

The modern versions of TB Hamiltonians include more involved scaling laws for the matrix elements, in an attempt to describe bonding of materials in various coordination environments (given by the number of near-neighboring atoms to a given atomic site) [95–98]. Effective many-body screening terms [96] and also self-consistent charge transfer have been included in some formulations [97]. These modern and more sophisticated versions are usually derived from databases of reference DFT calculations and hence have been termed DFT-TB. A high degree of accuracy is possible using these DFT-TB Hamiltonians, if the bonding patterns present in the atomic structure of the system do not depart too radically from those included in the fitting database.

Solving for the TBTE energies and forces requires diagonalization of the TB Hamiltonian, a procedure with a computational cost that scales as N_a^3 (N_a is the number of atoms in the system). Linear scaling or order-N [$\mathcal{O}(N)$] methods have been developed to overcome this bottleneck [99–101]. In $\mathcal{O}(N)$ methods, appropriate functional forms for the total electronic energy (the sum of eigenvalues in Eq. 7-2) are used, often based on a density-matrix representation of the occupied electronic sub-space. In the $\mathcal{O}(N)$ formulation of [100, 101], the sum of eigenvalues is written:

$$\begin{aligned}\sum_{i}^{N_\mathrm{occ}}\varepsilon_i &= Tr\left[\tilde{\rho}H\right]\ ,\\ &= Tr\left[\left(3\rho^2-2\rho^3\right)H\right]\ ;\end{aligned} \tag{7-4}$$

where H is the Hamiltonian matrix, and

$$\tilde{\rho}=3\rho^2-2\rho^3\ , \tag{7-5}$$

is the physical density matrix. The matrix ρ contains the electronic degrees of freedom of the problem. Unconstrained direct minimization of Eq. (7-4), with respect to matrix elements of ρ, leads to the minimum of the electronic energy, because, by construction, this functional form has a local minimum at the physical density matrix, i.e., the one corresponding to filling the N_occ lowest-energy eigenstates.

Truncation of the matrix elements in ρ within a given localized region surrounding each atomic site leads to the linear scaling or $\mathcal{O}(N)$ behavior of the method. This $\mathcal{O}(N)$ TBTE framework retains a quantum mechanical description of the system, at a quantitative level, when combined with the more modern TB Hamiltonian parameterizations [95–98].

In what follows, we discuss the main results of two recent investigations, one that employed classical potentials to address the scaling laws of the energies of SiNWs, and another that examines structural transitions of very thin SiNWs, using a combination of ab initio and TB methodologies.

7.3. STRUCTURAL PROPERTIES: APPLICATION OF EMPIRICAL METHODS

A major challenge in growing SiNWs is to control their final properties, such as the growth direction, surface types, reconstruction and passivation. Recently, it has been shown that the electronic properties of SiNWs are strongly dependent on their surface parameters [80], such that they can present metallic or insulating character. In order to grow SiNWs with pre-determined parameters, such as the growth direction and surface types, it is important to know the thermodynamic conditions which would favor growth with such properties. Additionally, it is crucial to know how the wire properties scale with the size of the nanowire.

In order to establish scaling laws for nanowires, a classical model has been used to determine the nanowire free energy in terms of its parameters [102, 103]. This energy could be used to establish the thermodynamic conditions for nanowire growth in equilibrium. The wire energy (F) comprises three elements: a bulk (E_b), a surface (E_s), and an edge (E_e) term,

$$F = E_\mathrm{e} + E_\mathrm{s} + E_\mathrm{b}\,. \tag{7-6}$$

The surface term is given by the contribution of all wire facets:

$$E_\mathrm{s} = \sum_i \gamma_i s_i\,, \tag{7-7}$$

where γ_i is the surface energy of facet i, and s_i is the number of unit cells of type i in the surface. Therefore, the nanowire energy lies between two limits [104]:

$$E_\mathrm{e} + \gamma_\mathrm{min} \sum_i s_i < (F - E_\mathrm{b}) < E_\mathrm{e} + \gamma_\mathrm{max} \sum_i s_i\,, \tag{7-8}$$

where γ_min and γ_max represent the minimum and maximum values for the surface energies, respectively. Therefore, the nanowire energies in a certain growth direction, with surfaces of mixed characters, should lie between those two limits. However, this modeling can only establish these limits but still does not allow to compare energies with different surface compositions [103].

The scaling properties of nanowires, such as their energies, have been described in terms of their diameters [80, 105, 106]. However, defining nanowire properties within a single parameter, such as diameter, is not simple, since nanostructures based on covalent bonding generally have facets and do not have a single diameter parameter. Authors either avoid defining such a parameter [69, 80] or describe the wire representative dimension as its smallest diameter, taken from images of the wire cross-section [106]. Other authors take the diameter of the smallest cylinder that contains the wire [46, 105]. If a single diameter is used, scaling of the nanowire energetics as function of its cross-section size takes into account an average of the surface energies of the various facets.

Considering the SiNW faceting, the wire perimeter (P), and not the wire diameter, provides a more detailed description of the nanowire scaling properties [104]. The wire perimeter comprises the sum of the length of each facet (f_i) of the wire ($P = \sum f_i$), shown in Figure 7-1. The surface size of each facet is determined by $f_i \times L$ (L is the wire length) and the total wire surface is given by $P \times L$. Therefore, a wire scaling law described in terms of its perimeter is equivalent to a law in terms of its total surface. Considering that in a nanowire, the surface/volume ratio is very large, it is reasonable to consider that scaling laws should be described in terms of the wire surface.

The wire energy can be given in terms of the facet length, f_i, shown in Figure 7-1. The energy limits of Eq. (7-8), can be written in terms of the wire perimeter ($P = \sum_i f_i$). Dividing all the terms in Eq. (7-8) by the number of atoms, $N \propto P^2$, per unit length, one gets the following relation for the wire energy per atom [104]:

$$E_e P^{-2} + \gamma_{min} P^{-1} < (F - E_b)/N < E_e P^{-2} + \gamma_{max} P^{-1}. \tag{7-9}$$

Equation (7-9) gives the limits for wire energy in terms of their perimeters. For large perimeters, the edge effects could be neglected, and the wire energies should have a linear relation with P^{-1}, and lie between two limiting cases, related to high (γ_{max}) and low (γ_{min}) surface energies. This model could be confirmed by atomistic simulations, combining interatomic potentials [91], discussed in Section 7.2.1, and molecular dynamics. Considering SiNWs in a ⟨001⟩ growth direction, one could have wires with different facets, ranging from wires with pure {100} surfaces to wires with pure {110} surfaces, as represented in Figure 7-2a.

Figure 7-2b shows the energy per atom (E_{nw}) of ⟨001⟩ SiNWs as a function of P^{-1}. This energy is defined with relation to the reference crystalline energy per

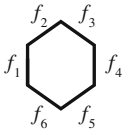


Figure 7-1. Cross-section of a typical silicon nanowire. The wire perimeter (P) is given as the sum of all the facet sides

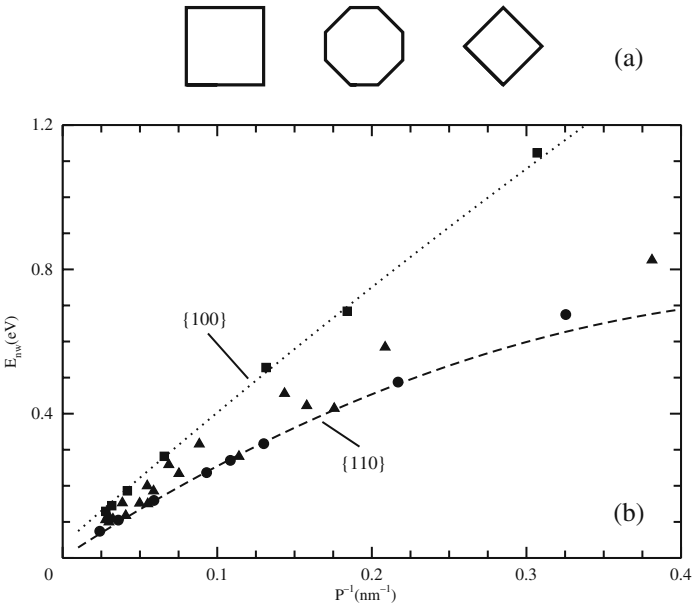


Figure 7-2. (**a**) A cross-sectional representation of the wires with ⟨001⟩ growth direction with pure {100}, mixed ({100} + {110}), and pure {110} surfaces. (**b**) Nanowire energy per atom (E_{nw}), computed using an empirical potential [91], as a function of P^{-1} for ⟨100⟩ nanowires. *Circles* (●) represent wires with pure {110} surfaces, the *squares* (■) those with pure {100} surfaces, and the *triangles* (▲) those with mixed ({100}+{110}) character. The *dotted* and *dashed lines* are data fittings coming from configurations that determine the energy limits

atom, such that very large wires ($P^{-1} \rightarrow 0$) tend to crystalline silicon, and the respective energy tends to zero. The nanowire energies follow a universal scaling law, for each facet family. The energy of a nanowire with any surface composition (pure or mixed character) falls within a certain region of the graphics, always between wires with {100} and {110} pure surfaces. These results are consistent: for a certain wire perimeter, the wire energy could have several values, depending on the surface types. The crystalline Si {100} surfaces have higher energies than {110} surfaces [107], therefore it is consistent that wires (with the same perimeter) have higher energies if they have {100} rather than {110} pure surfaces. Such results are fully consistent with the analytical model of Eq. (7-9). For large wire perimeters, edge effects can be neglected, and there is a linear relation between energy and the inverse of the wire perimeter. However, for smaller perimeters, edge effects become important [103, 108], leading to non-linear behavior. This behavior, of a window in which nanowire energies lie, could also be observed for SiNWs with ⟨110⟩, and ⟨112⟩ growth directions [104].

The results on Figure 7-2 provide the thermodynamic conditions for growing ⟨100⟩ SiNWs. Those results showed that wire properties are strongly dependent on the surface properties. However, another important aspect of these thin wires

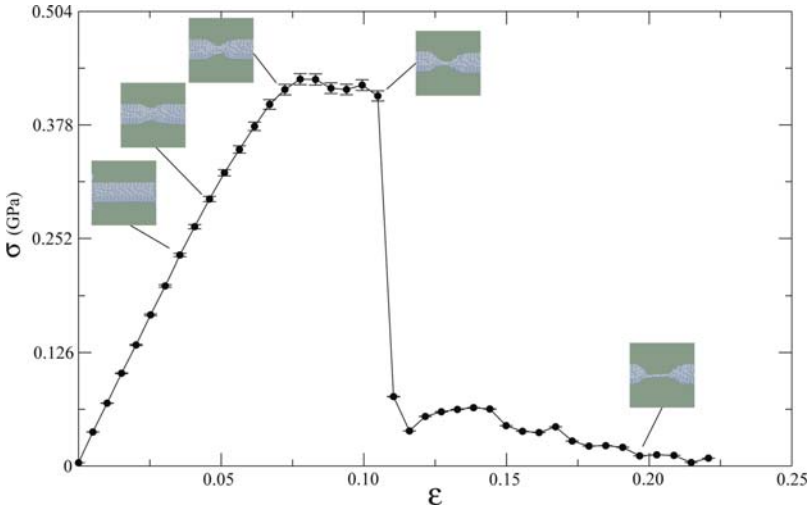


Figure 7-3. Stress (σ) strain (ε) curve of a thin wire ($\langle 001 \rangle$ SiNW with mixed $\{100\} + \{110\}$ surface character) computed using an empirical potential [91] and molecular dynamics at finite temperature ($T = 350\,\mathrm{K}$). The *inserts* represent the respective atomic wire configurations (near the deformation region) for several strains

is how they deform in response to external load, and the role of surfaces in such deformation processes. The typical behavior of a bulk material under load is a linear stress-strain relation for small strains, and for larger strains the material presents a stress saturation until rupture. Figure 7-3 shows the response of a SiNW to external load, computed combining an empirical potential and molecular dynamics. The wire response is essentially equivalent to that of a bulk material, but the wire appears to support considerably larger strains before full rupture. For small strains ($\varepsilon < 0.05$), the stress-strain curve has a linear behavior, which indicates an elastic response. For larger deformations ($0.05 < \varepsilon < 0.13$), inelastic behavior takes place. At about $\varepsilon \approx 0.13$, there is a large decrease in the stress, and the wire starts to open a surface crack. It would be expected that, for a certain large strain (here $\varepsilon \approx 0.13$), the nanowire would simply follow a fracture process, as observed in bulk experiments [106]. However, the evolution of the nanowire deformation allows larger strains, with considerably larger nanowire elongations. For those large strains, the crack did not evolve because atoms in the surface had enough thermal energy to diffuse toward the crack, preventing crack propagation and full rupture.

The above results show how faceting can be described by a scaling law that considers the nanowire perimeter and takes into account various possible surface terminations (or facets) for a given family of nanowires. In the following, we discuss a related but different issue, which is the stability of SiNWs, derived from bulk structures other than the cubic-diamond one, due to surface effects, for wires of very small diameters, of the order of 1 nm.

7.4. MORPHOLOGY OF THIN SILICON NANOWIRES: APPLICATION OF TIGHT BINDING AND FIRST PRINCIPLES METHODS

In the limit of the ultrathin wires, with diameters of the order of 2 nm and smaller, the surface-to-volume ratio is large enough that structural stability is strongly dictated by surface effects. Kagimura et al. [105] have investigated the possibility of structural transitions from the cubic-diamond (cd) nanowire forms to alternative geometries, for nanowires with diameters (D) between 0.4 and 2.0 nm. Calculations in [105] were performed in the DFT framework [83], within the generalized-gradient approximation (GGA) [109] and norm-conserving pseudopotentials [110–112], using the SIESTA code [113]. This ab initio methodology was combined with the $\mathcal{O}(N)$ density-matrix tight-binding methodology (DM-TBTE) of [100], using the TB silicon Hamiltonian of [95]. Being a linear-scaling method where the workload to compute the electronic structure of the system scales linearly with the number of atoms, as opposed to the cubic scaling of matrix-diagonalization methods, the DM-TBTE method is particularly useful in the study of larger structures not amenable to calculation by ab initio methods.

The sequence of phase transitions in the bulk [114–119], motivates most of the choice of geometries considered in [105]. They can be separated in three classes shown in Figures 7-4 and 7-5:

(1) *Diamond-structure nanowires* – These are derived from the cubic-diamond (cd) bulk phase, with the nanowire axis oriented along the [100] and [110] directions. The latter is the observed orientation of SiNWs with diameters between 3 and 10 nm [47]. The wires with $D > 1$ nm, oriented along [110], remain cd-like after the *ab initio* geometry optimization. Two examples (labelled cd1 and cd2) are shown in Figure 7-4a and b, respectively. Both wires undergo reconstruction at the surface but retain a crystalline core at the central interstitial channel. The cd wires with $D < 1$nm undergo extensive reconstruction towards amorphous-like structures.

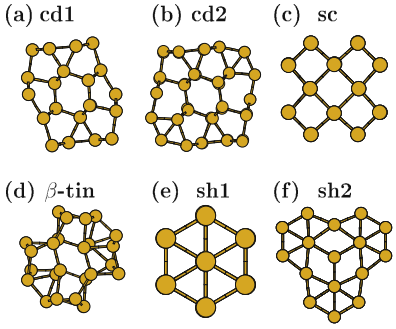


Figure 7-4. Cross sections of selected SiNW structures labeled according to the parent bulk phase. In (**a**) and (**b**), cd1 and cd2 wires derived from the cubic diamond structure, with axis along the [110] direction; in (**c**), a simple cubic wire; in (**d**), a $\beta-$tin wire with axis along the bulk c-direction; in (**e**) and (**f**), simple hexagonal wires with axis along the bulk c-direction (From [105])

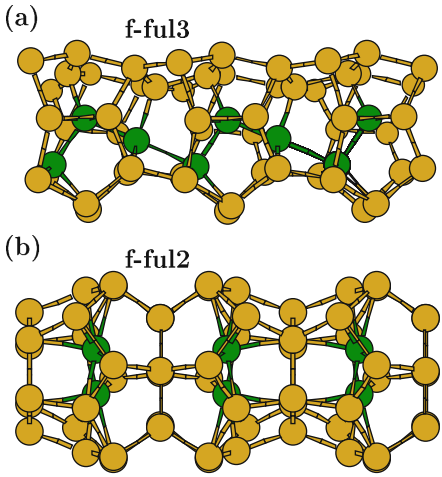


Figure 7-5. Side view of corrugated Si nanowire structures. In (**a**), corrugated wire resulting from a structural instability of a [100] cubic diamond nanowire; in (**b**), a filled-fullerene nanowire. Inner atoms are shown as *green spheres* (From [105])

(2) *High-density nanowires* – Derived from the high-density β-tin, simple cubic (sc), and simple hexagonal (sh) bulk phases. The simple cubic wire oriented along the [100] direction is shown in Figure 7-4c. It shows very little distortions relative to the bulk structure, and its energy is lower than that of the sc structures oriented along [110] and [111] directions. The relaxed geometry of the β-tin is shown in Figure 7-4d, with the nanowire axis parallel to the bulk *c*-axis, passing through the center of an interstitial channel. All sh nanowires oriented along the bulk *c* direction retain the crystalline order along the wire axis after geometry optimization, regardless of the wire radius. Two examples (sh1 and sh2) are shown in Figure 7-4e and f, respectively. An empty-hexagon variation of sh1 was also considered, where the atom at the hexagon center is removed.

(3) *Fullerene-like nanowires* – The structures in this class are derived from fullerene-like geometries [31, 67, 68]. Two of these are based, respectively, on the two fullerene-based geometries proposed in [31], namely, the Si_{20} cage polymer (ful1) and the Si_{24} cage polymer (ful2). Given the predicted stability of filled-fullerene-like clusters [67, 68], two variations of these nanowires were also considered, labeled f-ful1 and f-ful2, where two extra atoms are included inside the cage. The structure f-ful2 is shown in Figure 7-5b. Its corrugated structure, with the presence of fivefold rings at the surface, is similar to that of a wire which results from the reconstruction of the cd(100) wire, shown in Figure 7-5a. Hence, the latter was classified as fullerene-like in [105], and labeled f-ful3. Filled fullerene-like nanowires of smaller diameters, based on Si_{12} and Si_{16} cages, with one additional atom in the center of the cage, were also considered, but these have relatively high formation energies, as shown below.

Table 7-1. Calculated total energies per atom $\Delta E_{\rm tot}$, in eV/atom, of selected Si phases, relative to the cubic diamond phase. (From [105])

hd	β-tin	sh	sc	bcc	hcp	fcc
0.01	0.31	0.33	0.36	0.52	0.52	0.55

Table 7-1 shows the total energy per atom of the relevant bulk phases, $\Delta E_{tot} = E_{\rm tot} - E_{\rm tot}^{\rm cd}$, relative to the total energy of the cd phase ($E_{\rm tot}^{\rm cd}$). $\Delta E_{\rm tot}$ is within 0.20–0.40 eV/atom for the sc, sh, and β-tin phases, in good agreement with other calculations [114–116]. The calculated results for the diamond to β-tin transition pressure for Si (109 kbar) are also in good agreement with experimental results and other calculations [114–116].

Figure 7-6 shows the calculated total energies, $\Delta E_{\rm tot}$ (relative to the cd bulk phase, as defined previously) as a function of D, of the Si nanowires with $D <$ 1.4 nm. In the 0.9 nm $< D <$ 1.4 nm range, the formation energies of the high-density sc, sh, and β-tin nanowires, and also of the fullerene-like wires, are very close to the energies of cd1 and cd2, with energy differences of ∼0.05 eV/atom or less. These values are one order of magnitude smaller than the energy differences of the corresponding bulk phases in Table 7-1, showing that the energetics of wire formation, at such small diameters, is strongly affected by surface effects, consistent with findings discussed in Section 7.3. Figure 7-6 also shows that the amorphous wires, derived from the instabilities of thin cd wires, have higher formation energies than the high-density and the fullerene-like wires of comparable diameters. This suggests that amorphous wires in this diameter range could only be produced in conditions far from thermodynamic equilibrium.

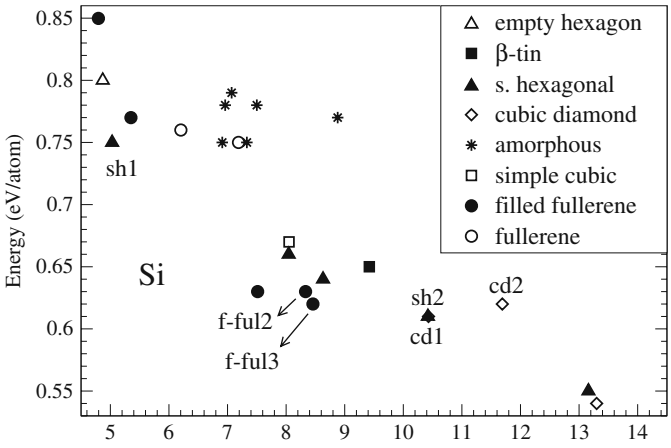


Figure 7-6. First-principles total energies (in eV/atom), relative to cd bulk energy, of Si nanowires as a function of nanowire diameter. Labeling of the structures is explained in the text (From [105])

A comment is in order regarding the hollow geometries like the unfilled-fullerene [69, 71, 75] and the empty-hexagon structures, to the extent that they have very high energies, when compared to the denser structures. Figure 7-6 also shows that among the nanowires with $D < 0.7$ nm, the sh1 structure appears below and to the left in the energy vs. diameter diagram, suggesting a high stability for this structure, when compared to the other small-diameter geometries. In the range $0.7\,\text{nm} < D < 0.9\,\text{nm}$, filled-fullerene-like wires are the most stable ones.

The results above indicate that the energetics of nanowire formation is determined by the interplay between the energies of a bulk part and a surface part of the wire. Figure 7-7 shows $-\Delta E_{\text{tot}}$ for the cd and sh Si wires, respectively, as a function of the inverse nanowire diameter D^{-1}. The values of the respective bulk phases (the $D^{-1} \rightarrow 0$ limit), are also shown. The Figure shows distinct trends for the cd- and sh-based Si nanowires, with a much larger variation of the total energy per atom as a function of D^{-1} for the cd-based wires than for the sh-based ones. Moreover, the energies of the two types of Si nanowires are very close for $D \sim 1.2$ nm. The results of the first-principles calculations shown in Figure 7-7 are reasonably well fitted by the expression

$$\varepsilon_{\text{nw}} = \varepsilon_{\text{s}} + (\varepsilon_{\text{b}} - \varepsilon_{\text{s}})\frac{(D - 2\rho^{-1/3})^2}{D^2} \tag{7-10}$$

In Eq. (7-10), ε_{nw} is ΔE_{tot} for the nanowire based on a given structure (cd or sh), while ε_b and ρ are, respectively, ΔE_{tot} and the number of atoms per volume for the corresponding bulk structure. ε_s is a measure of the surface energy per atom of the nanowire. ε_b and ρ are obtained from the bulk first-principles calculations, leaving

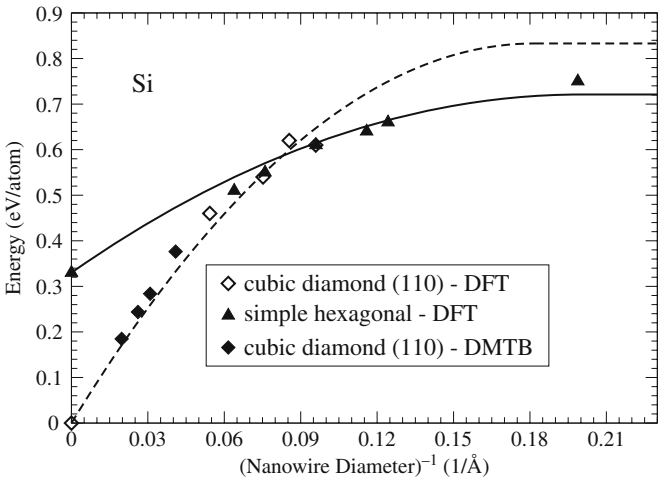


Figure 7-7. First-principles and tight-binding (DM-TBTE) total energies (in eV/atom), relative to the cd bulk energy, of SiNWs as a function of D^{-1}. The *dashed* (*solid*) *line* shows the curve obtained from a continuum model, parameterized for cd (sh) wires (From [105])

ε_s as the only fitting parameter. In Figure 7-7, Eq. (7-10) is plotted with $\varepsilon_s = 0.83\,\text{eV}$ for the cd structure and $\varepsilon_s = 0.72\,\text{eV}$ for the sh structure.

Equation (7-10) results from a simple continuum model for a cylindrical wire with diameter D, density ρ, and total energy E_{nw}. E_{nw} is decomposed into contributions due to the bulk-like atoms and to the low-coordinated surface atoms:

$$E_{\text{nw}} = \varepsilon_b \rho V_b + \varepsilon_s \rho V_s. \tag{7-11}$$

By considering a surface thickness of $\rho^{-1/3}$, it is straightforward to obtain Eq. (7-10).

Although this continuum model would be valid only in the limit $D^{-1} \to 0$, it provides a simple interpretation for the larger variation of ε_{nw} as a function of D^{-1} for the cd-based wires when compared with the sh-based ones. In Eq. (7-10), the variation of ε_{nw} with D is proportional to $\Delta\varepsilon = \varepsilon_{\text{s}} - \varepsilon_{\text{b}}$. $\Delta\varepsilon$ is about twice as large for the cd phase than for the sh phase, meaning that the energy cost of a cd surface is much larger than that of an sh surface. This arises from the fact that surface atoms in a cd structure are under-coordinated, while the surface atoms in an sh structure are still highly coordinated, reducing the energy cost of its surface. The above one-parameter continuum model, least-square-fitted to cd and sh geometries, displays a crossing between the two curves for $D \sim 1.2\,\text{nm}$. Hence, the model suggests a stability inversion resulting essentially from the larger $\Delta\varepsilon$ for the cd phase. The same stability inversion is indicated by the very close *ab initio* values obtained for ε_{nw} for the sh- and cd-based wires, for $D \sim 1.2\,\text{nm}$. Moreover, the structural instability of the cd class of nanowires suggested by the first principles results, from crystalline-like (for $D > 1\,\text{nm}$) to amorphous-like (for $D < 1\,\text{nm}$), occurs at diameters which are very near the cd-sh stability inversion. DM-TBTE calculations allows one to verify the validity of Eq. (7-10) to wires of larger radii. The results for (110)-oriented Si wires with $2\,\text{nm} > D > 5\,\text{nm}$, which were not used in the fitting of Eq. (7-10), are also included in Figure 7-7. The agreement between the *ab initio*-fitted Eq. (7-10) and the DM-TBTE values for ε_{nw} is very good, at larger diameters.

7.5. CONCLUSIONS

Modern theoretical calculations have reached a point where realistic description of materials properties are provided by computational simulations. A carefully constructed empirical potential, such as EDIP, provides a good description of silicon-nanowire energetics. When combined with a scaling law that considers the nanowire perimeter, it allows for the comparison of silicon wires with different surface terminations (or facets) for a given family of nanowires. Further, simulations using this potential suggest that the response under external load of silicon nanowires is different from the bulk, to the extent that crack propagation is suppressed and healing at the crack is mediated by surface effects. Investigation of nanowire properties are also within the reach of tight-binding and first principles methodologies. These methods have been used to describe how, in the limit of very

thin wires, of subnanometer diameters, surface effects may lead to the stabilization of nanowire structures built from the simple-hexagonal bulk structure or from filled-fullerene units.

REFERENCES

1. Campaño R, Molenkamp L, Paul DJ (eds) (1997) Technology Roadmap for Nanoelectronics in "European Commision for Future and Emerging Technologies".
2. Morales AM, Lieber CM (1998) Science 279: 208.
3. Gudiksen MS, Lauhon LJ, Wang J, Smith DC, Lieber CM (2002) Lieber Nature 415: 617.
4. Melosh NA, Boukai A, Diana F, Gerardot B, Badolato A, Petroff PM, Heath JR (2003) Science 300: 112.
5. Xia YN, Yang PD, Sun YG, Wu YY, Mayers B, Gates B, Yin YD, Kim F, Yan YQ (2003) Adv Mater 15: 353.
6. Lu W, Lieber CM (2007) Nat Mater 6: 841.
7. Stone NJ, Ahmed H (1998) Appl Phys Lett 73: 2134.
8. Stone NJ, Ahmed H (1998) Microeletr Eng 42: 511.
9. Cui Y, Lieber CM (2001) Science 291: 851.
10. Huang Y, Duan XF, Cui Y, Lauhon LJ, Kim KH, Lieber CM (2001) Science 294: 1313.
11. Chung SW, Yu JY, Heath JR (2000) Appl Phys Lett 76: 2068.
12. Duan XF, Huang Y, Agarwal R, Lieber CM (2003) Nature 421: 241.
13. Huang MH, Mao S, Feick H, Yan HQ, Wu YY, Kind H, Weber E, Russo R, Yang PD (2001) Science 292: 1897.
14. Cui Y, Wei QQ, Park HK, Lieber CM (2001) Science 293: 1289.
15. Zhou XT, Hu JQ, Li CP, Ma DDD, Lee CS, Lee ST (2003) Chem Phys Lett 369: 220.
16. Patolsky F, Zheng GF, Lieber CM (2006) Nat Protoc 1: 1711.
17. Iijima S, Ichihashi T (1993) Nature 363: 603.
18. Cui Y, Zhong ZH, Wang DL, Wang WU, Lieber CM (2003) Nano Lett 3: 149.
19. Lehmann V, Gosele U (1991) Appl Phys Lett 58: 856.
20. Stucky GD, Macdougall JE (1990) Science 247: 669.
21. Ridley BK (1991) Rep Prog Phys 54: 169.
22. Liu HI, Biegelsen DK, Ponce FA, Johnson NM, Pease RFW (1994) Appl Phys Lett 64: 1383.
23. Liu HI, Biegelsen DK, Johnson NM, Ponce FA, Pease RFW (1994) J Vac Sci Tech 11: 2532.
24. Namatsu H, Takahashi Y, Nagase M, Murase K (1995) J Vac Sci Tech B 13: 2166.
25. Kurihara K, Iwadate K, Namatsu H, Nagase M, Takenaka H, Murase K (1995) Jn J Appl Phys 34: 6940.
26. Ono T, Saitoh H, Esashi M (1997) Appl Phys Lett 70: 1852.
27. Almawlawi D, Liu CZ, Moskovits M (1994) J Mater Res 9: 1014.
28. Coleman NRB, Morris MA, Spalding TR, Holmes JD (2001) J Am Chem Soc 123: 187.
29. Zhang YF, Tang YH, Wang N, Yu DP, Lee CS, Bello I, Lee ST (1998) Appl Phys Lett 72: 1835.
30. Wang N, Tang YH, Zhang YF, Lee CS, Lee ST (1998) Phys Rev B 58: 16024.
31. Marsen B, Sattler K (1999) Phys Rev B 60: 11593.
32. Yu DP, Bai ZG, Ding Y, Hang QL, Zhang HZ, Wang JJ, Zou YH, Qian W, Xiong GC, Zhou HT, Feng SQ (1998) Appl Phys Lett 72: 3458.
33. Gole JL, Stout JD, Rauch WL, Wang ZL (2000) Appl Phys Lett 76: 2346.
34. Zhang YF, Tang YH, Lam C, Wang N, Lee CS, Bello I, Lee ST (2000) J Cryst Growth 212: 115.
35. Pan ZW, Dai ZR, Xu L, Lee ST, Wang ZL (2001) J Phys Chem 105: 2507.

36. Westwater J, Gosain DP, Tomiya S, Usui S, Ruda H (1997) J Vacc Sci Tech 15: 554.
37. Holmes JD, Johnston KP, Doty RC, Korgel BA (2000) Science 287: 1471.
38. Gudiksen MS, Lieber CM (2000) J Am Chem Soc 122: 8801.
39. Kamins TI, Williams RS, Chen Y, Chang YL, Chang YA (2000) Appl Phys Lett 76: 562.
40. Zhang RQ, Lifshitz Y, Lee ST (2003) Adv Mater 15: 635.
41. Lu XM, Hanrath T, Johnston KP, Korgel BA (2003) Nano Lett 3: 93.
42. Hanrath T, Korgel BA (2003) Adv Mater 15: 437.
43. Hofmann S, Sharma R, Wirth CT, Cervantes-Sodi F, Ducati C, Kasama T, Dunin-Borkowski RE, Drucker J, Bennett P, Robertson J (2008) Nat Mater 7: 372.
44. Kodambaka S, Tersoff J, Reuter MC, Ross FM (2006) Phys Rev Lett 96: 096105.
45. Ross FM, Tersoff J, Reuter MC (2005) Phys Rev Lett 95: 146104.
46. Ma DDD, Lee CS, Au FCK, Tong SY, Lee ST (2003) Science 299: 1874.
47. Wu Y, Cui Y, Huynh L, Barrelet CJ, Bell DC, Lieber CM (2004) Nano Lett 4: 433.
48. Jarrold MF, Constant VA (1991) Phys Rev Lett 67: 2994.
49. Hunter JM, Fye JL, Jarrold MF, Bower JE (1994) Phys Rev Lett 73: 2063.
50. Menon M, Richter E (1999) Phys Rev Lett 83: 792.
51. Fagan SB, Baierle RJ, Mota R, da Silva AJR, Fazzio A (2000) Phys Rev B 61: 9994.
52. Zhang RQ, Lee ST, Law CK, Li WK, Teo BK (2002) Chem Phys Lett 364: 251.
53. Barnard AS, Russo SP (2003) J Phys Chem 107: 7577.
54. Zhang M, Kan YH, Zang OJ, Su ZM, Wang RS (2003) Chem Phys Lett 379: 81.
55. Zhang RQ, Lee HL, Li WK, Teo BK (2005) J Phys Chem 109: 8605.
56. Pradhan P, Ray AK, (2006) J Comput Theor Nanosc 3: 128.
57. Kang JW, Seo JJ, Hwang HJ (2002) J Nanosci Nanotech 2: 687.
58. Dumitrică T, Hua M, Yakobson BI (2004) Phys Rev B 70: 241303.
59. Zhao MW, Zhang RQ, Xia YY, Song C, Lee ST (2007) J Phys Chem C 111: 1234.
60. Yan BH, Zhou G, Wu J, Duan WH, Gu BL (2006) Phys Rev B 73: 155432.
61. Yan BH, Zhou G, Zeng XC, Wu J, Gu BL, Duan WH (2007) Appl Phys Lett 91: 103107.
62. Ni M, Luo GF, Lu J, Lai L, Wang L, Jing MW, Song W, Gao ZX, Li GP, Mei WN, Yu DP (2007) Nanotechnology 18: 505707.
63. Zhang DB, Hua M, Dumitrică T (2008) J Chem Phys 128: 084104.
64. Li BX, Cao PL, Zhang RQ, Lee ST (2002) Phys Rev B 65: 125305.
65. Zhao YF, Yakobson BI (2003) Phys Rev Lett 91: 035501.
66. Zhang RQ, Lifshitz Y, Ma DDD, Zhao YL, Frauenheim T, Lee ST, Tong SY (2005) J Chem Phys 123: 144703.
67. Räthlisberger U, Andreoni W, Parrinello M (1994) Phys Rev Lett 72: 665.
68. Landman U, Barnett RN, Scherbakov AG, Avouris P (2000) Phys Rev Lett 85: 1958.
69. Ponomareva I, Menon M, Srivastava D, Andriotis AN (2005) Phys Rev Lett 95: 265502.
70. Akiyama T, Nakamura K, Ito T (2006) Phys Rev B 74: 033307.
71. Ponomareva I, Menon M, Richter E, Andriotis AN (2006) Phys Rev B 74: 125311.
72. Liu SD, Jayanthi CS, Zhang ZY, Wu SY (2007) J Comput Theor Nanosci 4: 275.
73. Maeda S, Akiyama T, Nakamura K, Ito T (2007) J Cryst Growth 301: 871.
74. Ng MF, Zhou LP, Yang SW, Sim LY, Tan VBC, Wu P (2007) Phys Rev B 76: 155435.
75. Ponomareva I, Richter E, Andriotis AN, Menon M (2007) Nano Lett 7: 3424.
76. Lu N, Ciobanu CV, Chan TL, Chuang FC, Wang CZ, Ho KM (2007) J Phys Chem 111: 7933.
77. Sorokin PB, Avramov PV, Kvashnin AG, Kvashnin DG, Ovchinnikov SG, Fedorov AS (2008) Phys Rev B 77: 235417.
78. Ma L, Wang HG, Zhao JJ, Wang GH (2008) Chem Phys Lett 452: 183.
79. Rurali R, Cartoixa X, Galvao DS (2008) Phys Rev B 77: 073403.
80. Rurali R, Lorente N (2005) Phys Rev Lett 94: 026805.

81. Wu ZG, Neaton JB, Grossman JC (2008) Phys Rev Lett 100: 246804.
82. Hohenberg P, Kohn W (1964) Phys Rev 136: B864.
83. Kohn W, Sham LJ (1965) Phys Rev 140: A1133.
84. Payne MC, Teter MP, Allan DC, Arias TA, Joannopoulos JD (1992) Rev Mod Phys 64: 1045.
85. Nguyen-Manh D, Vitek V, Horsfield AP (2007) Prog Mater Sci 52: 255.
86. Beck TL (2000) Rev Mod Phys 72: 1041.
87. Balamane H, Halicioglu T, Tiller WA (1992) Phys Rev B 46: 2250.
88. Justo JF (2005) In: S. Yip (ed) Handbook of Materials Modeling, Springer, Dordrecht.
89. Stillinger FH, Weber TA (1985) Phys Rev B 31: 5262.
90. Tersoff J (1988) Phys Rev B 37: 6991.
91. Justo JF, Bazant MZ, Kaxiras E, Bulatov VV, Yip S (1998) Phys Rev B 58: 2539.
92. Menezes RD, Justo JF, Assali LVC (2007) Phys Stat Solid A 204: 951.
93. Carlsson AE (1990) In Ehrenreich H, Turnbull D (eds) Solid State Physics, vol. 43 Academic Press, San Diego, pp. 1-91.
94. Harrison WA (1989) In: Electronic Structure and the Properties of Solids, Dover Publications, New York.
95. Kwon I, Biswas R, Wang CZ, Ho KM, Soukoulis CM (1994) Phys Rev B 49: 7242.
96. Tang MS, Wang CZ, Chan CT, Ho KM (1996) Phys Rev B 53: 979.
97. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G (1998) Phys Rev B 58: 7260.
98. Lenosky TJ, Kress JD, Kwon I, Voter AF, Edwards B, Richards DF, Yang S, Adams JB (1997) Phys Rev B 55: 1528.
99. Goedecker S (1999) Rev Mod Phys 71: 1085.
100. Li XP, Nunes RW, Vanderbilt D (1993) Phys Rev B 47: 10891.
101. Nunes RW, Vanderbilt D (1994) Phys Rev B 50: 17611.
102. Herring C (1951) Phys Rev 82: 87.
103. Zhao Y, Yakobson BI (2003) Phys Rev Lett 91: 035501.
104. Justo JF, Menezes RD, Assali LVC (2007) Phys Rev B 75: 045303.
105. Kagimura R, Nunes RW, Chacham H (2005) Phys Rev Lett 95: 115502.
106. Kizuka T, Takatani Y, Asaka K, Yoshizaki R (2005) Phys Rev B 72: 035333.
107. Stekolnikov AA, Bechstedt F (2005) Phys Rev B 72: 125326.
108. Ismail-Beigi S, Arias T (1998) Phys Rev B 57: 11923.
109. Perdew JP, Burke K, Ernzerhof M (1996) Phys Rev Lett 77: 11593.
110. Troullier N, Martins JL (1991) Phys Rev B 43: 1993.
111. Kleinman L, Bylander DM (1991) Phys Rev Lett 48: 1425.
112. Gonze X, Stumpf R, Scheffler M (1991) Phys Rev B 44: 8503.
113. Ordejón P, Artacho E, Soler JM (1996) Phys Rev B 53: R10441.
114. Moll N, Bockstedte M, Fuchs M, Pehlke E, Scheffler M (1995) Phys Rev B 52: 2550.
115. Lee IH, Martin RM (1997) Phys Rev B 56: 7197.
116. Cheng C (2003) Phys Rev B 67: 134109.
117. Mujica A, Rubio A, Munoz A, Needs RJ (2003) Rev Mod Phys 75: 863.
118. Gaál-Nagy K, Pavone P, Strauch D (2004) Phys Rev B 69: 134112.
119. Poswal HK, Garg N, Sharma SM, Busetto E, Sikka SK, Gundiah G, Deepak FL, Rao CNR (2005) J Nanosci Nanotech 5: 729.

CHAPTER 8

MULTISCALE MODELING OF SURFACE EFFECTS ON THE MECHANICAL BEHAVIOR AND PROPERTIES OF NANOWIRES

HAROLD S. PARK[1] AND PATRICK A. KLEIN[2]
[1] *University of Colorado at Boulder, Boulder, CO, USA,e-mail: harold.park@colorado.edu*
[2] *Franklin Templeton Investments, San Mateo, CA, USA*

Abstract: Surface effects have recently been recognized as having the dominant effect on the mechanical behavior and properties of nanowires. Understanding these effects will be critical, in particular for the accurate design and functionalization of future nanowire-based nanoelectromechanical systems, including sensors, resonators and actuators. The purpose of this chapter is therefore to overview a recently developed multiscale continuum model, the surface Cauchy-Born model, which was developed to study nanomaterials where surface effects such as surface stresses are expected to contribute significantly to the mechanical response. The approach is based upon a simple extension to Cauchy-Born theory, in which continuum properties such as stress and stiffness are obtained for a given material and crystal structure directly from an underlying atomistic potential. In particular, by explicitly accounting for differences in energy for both bulk and surface atoms, we develop a variational formulation that leads to a nanomechanical boundary value problem that can be solved using standard nonlinear finite element methods for displacements, stresses and strains while naturally accounting for the effects of atomistic surface stresses. Finite element calculations using the proposed surface Cauchy-Born model demonstrate how surface stresses cause variations in the resonant frequencies of silicon nanowires as compared to those expected from continuum beam theory, and emphasize the importance of nonlinear elasticity in understanding and capturing the resonant frequency variations

Keywords: Multiscale computations, Nanowires

8.1. INTRODUCTION

Nanowires have been amongst the most studied nanomaterials in recent years. The intense interest in nanowires has emerged for a variety of reasons, foremost because their small sizes often lead to unique physical properties that are not observed in the corresponding bulk material. Non-bulk phenomena have been observed in the mechanical, electrical, thermal, and optical properties of both metallic and

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 193–229.
DOI 10.1007/978-1-4020-9785-0_8,

semiconducting nanowires [1–3]. These unique properties have therefore generated significant interest in using nanowires as the basic building blocks of future multifunctional nanoelectromechanical systems (NEMS) [4, 5], which have been proposed for a multitude of cross-disciplinary applications, including chemical and biological sensing, force and pressure sensing, high frequency resonators, and many others [4–9].[1,2]

The potential of nanowires in future nanotechnologies has led to significant interest in experimental characterization of the size-dependent elastic properties of nanowires. The experimental techniques utilized have varied from time-resolved spectroscopy [10] to AFM-induced bending [11–12] or resonance measurements [6, 22–30]. In general, resonance measurements to obtain the nanoscale elastic properties are predominant in the literature due to their relative simplicity as compared to bending and tensile experiments at the nanoscale due to the reduced amount of nanowire manipulation involved in resonance-based testing. The experimental results show significant scatter, with some predictions of enhanced elastic stiffness [14, 15, 23], some predicting reduced elastic stiffness [10, 29, 31, 32] with decreasing nanostructure size, and some predicting no change with respect to the bulk elastic stiffness [12,13]. Because many of the proposed applications for nanowire-based NEMS, such as resonant mass sensing and high frequency oscillators [4, 5, 8] rely on the ability to control and tailor the nanowire resonant frequencies with a high degree of precision, it is critical to be able to predict how the elastic properties of nanowires scale with size.

In analyzing the mechanical behavior of nanomaterials, a key feature of interest is intrinsic surface stresses that arise due to their large ratio of surface area to volume [33]. Surface stresses have recently been found to cause phase transformations in gold nanowires [34], self-healing behavior in metal nanowires [35–37], and surface reorientations in thin metallic films and wires [38, 39]. Surface and confinement effects are also known to cause elevated strength in nanomaterials [11,12,14,15], orientation-dependent surface elastic properties [40–42] and a first-order effect on the operant modes of inelastic deformation in metal nanowires [43].

The knowledge that surface effects are critical to understanding the mechanical behavior and properties of nanomaterials has motivated the development of enhanced continuum models, as standard continuum mechanics is length scale independent. Various analytic models have been developed to study the effects of surface stress on the resonant properties of nanobeams [44–50], or more generally to capture the non-bulk mechanical behavior and properties of nanostructures [41, 44–60]. Due to assumptions utilized to make the analyses tractable, the coupled effects of geometry, surface orientation and system size on the resonant properties of nanowires have not been quantified, nor have surface stress effects arising directly

---

[1] Portions of Section 1, 2.1, 2.2 and 3.1 are from [87]. Copyright John Wiley and Sons Limited. Reproduced with permission.

[2] Portions of Section (2.4), (6), (7) and (8) are reprinted with permission from [84]. Copyright (2008), American Institute of Physics.

from atomistic principles been included in the analyses, which are generally in two-dimensions. The analyses also utilize overly simplistic pair-type atomic interactions to describe the surface physics, which tend to incorrectly predict a compressive surface stress for metals whereas the surface stress for metals is almost always tensile. Furthermore, these models are based upon linear elastic continuum mechanics. These errors indicate that quantitative analyses for real materials cannot be made using these approaches.

Alternatively, multiple scale models of nanomaterials have been developed to combine the insights into the detailed response of materials that are available through atomistics with the reduced computational expense that continuum approaches offer. Methods for both quasistatic [61–64] and dynamic coupling of atomistics and continua [65–74] have been proposed. With few exceptions [64], a critical issue with these methods is that the continuum region generally surrounds or encloses the atomistic region, thereby eliminating the effects of surface stresses on the atomistic behavior.

Therefore, the purpose of this work is to overview a recently developed nonlinearly elastic, finite deformation multiscale continuum model that incorporates atomistic surface stress effects to study surface effects on the mechanical behavior and properties of nanomaterials. We accomplish this through a decomposition of the potential energy of the system into bulk and surface components; while this decomposition has been considered before [33, 51, 57], those works require either higher order terms in the surface energy or empirical fits to constants for the surface stress which require additional atomistic simulations. The uniqueness of the present approach is that the surface energies are obtained directly from an underlying crystal structure and interatomic potential; this approach is adopted since a direct link to the underlying atomic structure is desired for the constitutive response, and constitutes the multiscale nature of the approach. Therefore, the approach taken in this work uses much of the machinery typically used in Cauchy-Born constitutive modeling [61,75,76] with care taken to treat surface unit cells correctly. This modification to treat the surface unit cells differently is the key to utilizing the Cauchy-Born rule to model surface effects in nanostructures as the Cauchy-Born model is based upon a bulk atomic unit cell that observes no free surface effects.

By decomposing the potential energy into bulk and surface components, a variational formulation that is composed of surface and volumetric contributions to the potential energy is obtained. Thus, as the structural length scale decreases and the ratio of surface area to volume increases, the correct surface energy contribution to the overall system energy is naturally obtained. Because the method is based on an energetic approach, the solution of the variational equation can be readily obtained using standard nonlinear finite element techniques; as the finite element stresses are simply derivatives of the strain energy, the effects of the surface energies are transferred naturally to the numerical model. This fact constitutes a distinct advantage for the proposed approach as it can therefore be utilized to solve boundary value problems for the deformation of nanoscale materials with arbitrary geometries, surface orientations and external loading. We present finite element calculations verifying the accuracy of the proposed surface Cauchy-Born model as

compared to benchmark fully atomistic calculations for FCC metal nanowires, an extension of the surface Cauchy-Born for diamond cubic lattices and silicon, and also using the surface Cauchy-Born model to investigate how surface stress effects cause deviations in the resonant frequencies of silicon nanowires as compared to those expected from continuum beam theory.

8.2. METHODOLOGY

8.2.1. Continuum Mechanics Preliminaries

In this section, we briefly review some elements of nonlinear continuum mechanics which are central to the Cauchy-Born formulation. The position of a material point $\mathbf{X}$ in the reference configuration can be mapped to the current configuration $\mathbf{x}$ via

$$\mathbf{x} = \mathbf{X} + \mathbf{u}(\mathbf{X}), \tag{8-1}$$

where $\mathbf{u}(\mathbf{X})$ is the displacement. The transformation of an infinitesimal line segment from the reference to the current configuration is described by the deformation gradient $\mathbf{F}$, which is defined as

$$\mathbf{F} = \frac{\partial \mathbf{X}}{\partial \mathbf{X}} = \mathbf{I} + \frac{\partial \mathbf{u}}{\partial \mathbf{X}}, \tag{8-2}$$

where $\mathbf{I}$ is the identity tensor. In Green elastic theory, stress is derived by differentiating the material strain energy density function. In order to satisfy material frame indifference, the strain energy density must be expressed as a function of the right stretch tensor $\mathbf{C}$,

$$W(\mathbf{F}) = \boldsymbol{\Phi}(\mathbf{C}), \tag{8-3}$$

where

$$\mathbf{C} = \mathbf{F}^{\mathbf{T}}\mathbf{F}. \tag{8-4}$$

From the strain energy density, one can obtain the first ($\mathbf{P}$) and second ($\mathbf{S}$) Piola-Kirchoff stresses as

$$\mathbf{P} = \frac{\partial W(\mathbf{F})}{\partial \mathbf{F}} \quad \text{and} \quad \mathbf{S} = 2\frac{\partial \Phi(\mathbf{C})}{\partial \mathbf{C}}, \tag{8-5}$$

where the Piola-Kirchoff stresses are related by

$$\mathbf{P} = \mathbf{S}\mathbf{F}^{\mathbf{T}}. \tag{8-6}$$

For crystalline materials, we can construct a strain energy density function by considering the bonds in a representative volume of the crystal. For the case of

a centrosymmetric crystal modeled using only pair interactions, the strain energy density is defined in terms of the interatomic potential U as [75]

$$\Phi(\mathbf{C}) = \frac{1}{2}\frac{1}{\Omega_0^a}\sum_{i=1}^{n_b} U\left(r^{(i)}(\mathbf{C})\right). \tag{8-7}$$

In (8-7), n_b is the total number of bonds to a representative bulk atom, Ω_0^a is the representative atomic volume in the undeformed configuration, $r^{(i)}$ is the defomed bond length which follows the relationship

$$r^{(i)} = \sqrt{\mathbf{R}_0^{(i)} \cdot \mathbf{C}\mathbf{R}_0^{(i)}}, \tag{8-8}$$

where $\mathbf{R}_0$ is the undeformed bond vector, and the factor of $1/2$ in (8-7) comes from splitting the energy of each bond.

The strain energy density (8-7) is exact in describing the change in energy per volume of a bulk atom in a corresponding defect-free atomistic system subject to homogeneous deformation. From (8-5) and (8-7), the second Piola-Kirchoff stress is given by

$$\mathbf{S}(\mathbf{C}) = \frac{1}{\Omega_0^a}\sum_{i=1}^{n_b}\left(U'(r^{(i)})\frac{\partial r^{(i)}}{\partial \mathbf{C}}\right). \tag{8-9}$$

These assumptions constitute the Cauchy-Born hypothesis; we note that the Cauchy-Born model has been used previously for a variety of different materials and lattices, including carbon nanotubes [76,77] as well as semiconductors such as silicon [60,78,79] and FCC metals [61].

As mentioned above, all points at which the Cauchy-Born hypothesis is applied are assumed to lie in the bulk because $\Phi(\mathbf{C})$ does not account for surface effects. Therefore, the issue at hand is to develop an expression for the energy density along the surfaces of a body. We discuss an approach to accomplishing this next.

8.2.2. Surface and Bulk Energy Densities

In this section, we discuss the methodology by which the the total atomistic potential energy of a body is represented by continuum energy densities with appropriate representations for bulk and surface energy densities. The relationship between the continuum strain energy and the total potential energy of the corresponding, defect-free atomistic system can be approximated as

$$\sum_{\alpha=1}^{\text{natoms}} U_\alpha(r) \approx \int_{\Omega_0^{\text{bulk}}} \Phi(\mathbf{C})\, d\Omega + \int_{\Gamma_0} \gamma(\mathbf{C})\, d\Gamma, \tag{8-10}$$

where U_α is the potential energy of atom α, r is the interatomic distance, $\Phi(\mathbf{C})$ is the bulk strain energy density introduced in Section 8.2.1, Ω_0^{bulk} represents the

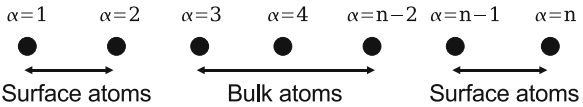


Figure 8-1. Illustration of bulk and surface atoms for a 1D atomic chain with second nearest neighbor interactions

volume of the body in which all atoms are fully coordinated, $\gamma(\mathbf{C})$ is the surface energy density, Γ_0 represents the surface area of the body in which the atoms are undercoordinated and *natoms* is the total number of atoms in the system.

Analogous to the bulk energy density, we will derive this surface energy density $\gamma(\mathbf{C})$ to describe the energy per representative undeformed *area* of atoms at or near the surface of a homogeneously deforming crystal. Figure 8-1 illustrates the decomposition given by (8-10). The bulk strain energy density function $\Phi(\mathbf{C})$ is integrated only over the part of the domain composed of fully coordinated atoms, or atoms $\alpha = 3 \rightarrow n-2$ in Figure 8-1. The potential energy of the atoms at or near the surface (atoms $\alpha = 1, 2, n-1, n$ in Figure 8-1) which do not possess a bulk bonding configuration is represented by the surface energy density $\gamma(\mathbf{C})$. In order to derive the surface energy density $\gamma(\mathbf{C})$ with a Cauchy-Born approach, we need to identify the *surface unit cell*, or the cluster of atoms that reproduce the structure of the surface layers when repeated in the plane of the surface. The surface unit cell possesses translational symmetry only in the plane of the surface, unlike the bulk unit cell which possesses translational symmetry in all directions. As illustrated in the Figure 8-1, each layer of atoms near the surface has a different bonding configuration. With these considerations, we express the surface energy density generally as

$$\gamma(\mathbf{C}) = \frac{1}{2}\frac{1}{\Gamma_0^a}\sum_{i=1}^{n_{\mathrm{sl}}}\sum_{j=1}^{n_{\mathrm{b}_i}} U(r^{(j)}(C)), \tag{8-11}$$

where n_{sl} is the number of surface layers, n_{b_i} is the number of bonds for atoms in surface layer i, Γ_0^a is the representative area of the entire surface layer cluster and the factor of $1/2$ again comes due to splitting the energy of each bond.

We can immediately define the surface stress resulting from the surface energy in (8-11) as

$$\tilde{\mathbf{S}}(\mathbf{C}) = 2\frac{\partial\gamma(\mathbf{C})}{\partial\mathbf{C}} = \frac{1}{\Gamma_0^a}\sum_{i=1}^{n_{\mathrm{sl}}}\sum_{j=1}^{n_{\mathrm{b}_i}}\left(U'(r^{(j)})\frac{\partial r^{(j)}}{\partial\mathbf{C}}\right). \tag{8-12}$$

Figure 8-2 summarizes the basic idea of the surface Cauchy-Born model. For both bulk and surface components, the underlying atomistic potential energy is obtained by subjecting the bulk or surface unit cell to the continuum stretch tensor $\mathbf{C}$. Once the strain energy density of the deformed unit cell is known, the

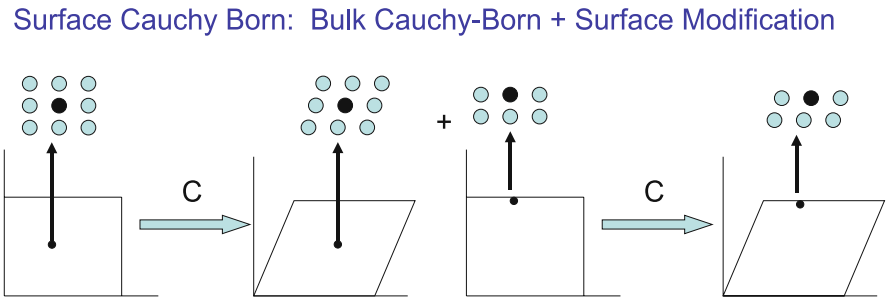


Figure 8-2. Illustration of difference between bulk and surface contributions to surface Cauchy-Born model

bulk stress can be calculated using (8-9), while the surface stress can be calculated using (8-12).

8.2.3. Formulation for Embedded Atom Method/FCC Metals

We now present the SCB formulation for FCC metals. We first note that FCC transition metals can be modeled with a high degree of fidelity including surface driven phenomena by the embedded atom method (EAM) [35,36,80]. Considering a purely atomistic system, the EAM energy for an atom U_i is written as[3]

$$\begin{aligned} U_i &= F_i(\bar{\rho}_i) + \frac{1}{2}\sum_{j\neq i}^{n\mathrm{b_i}} \phi_{ij}(r_{ij}), \\ \bar{\rho}_i &= \sum_{j\neq i}^{n\mathrm{b}_i} \rho_j(r_{ij}), \end{aligned} \tag{8-13}$$

where $n\mathrm{b}_i$ are the number of bonds of atom i, F_i is the embedding function, $\bar{\rho}_i$ is the total electron density at atom i, ρ_j is the contribution to the electron density at atom i from atom j, ϕ_{ij} is a pair interaction function and r_{ij} is the distance between atoms i and j. We note that the number of bonds $n\mathrm{b}_i$ is dictated by the cutoff distance of the interatomic potential.

In order to turn the atomistic potential energy into a form suitable for the CB approximation, two steps are taken. First, the potential energy is converted into a strain energy density through normalization by a representative atomic volume Ω_0; Ω_0 can be calculated noting that there are 4 atoms in an FCC unit cell of volume a_0^3, where a_0 is the lattice parameter. Thus $\Omega_0 = 4/a_0^3$ for a $\langle 100\rangle$ oriented crystal.

[3] Portions of Section 2.3, 3.1 and 5, including Figures 3,6,7,8, are reprinted with permission from [88]. Copyright (2007) by the American Physical Society.

Second, the neighborhood surrounding each atom is constrained to deform homogeneously via continuum mechanics quantities such as the deformation gradient $\mathbf{F}$, or the stretch tensor $\mathbf{C} = \mathbf{F}^T\mathbf{F}$. The resulting EAM strain energy density Φ is

$$\begin{aligned} \Phi(\mathbf{C}) &= \frac{1}{\Omega_0}\left(F_i(\bar{\rho}_i) + \phi_i\right), \\ \phi_i &= \frac{1}{2}\sum_{j\neq i}^{\mathrm{nbrv}_i} \phi_{ij}(r_{ij}(\mathbf{C})), \\ \bar{\rho}_i &= \sum_{j\neq i}^{\mathrm{nbrv}_i} \rho_j(r_{ij}(\mathbf{C})), \end{aligned} \tag{8-14}$$

where nbrv_i are the number of bonds in the representative unit volume Ω_0 for atom i, F_i is the embedding function, ρ_j is the contribution to the electron density at atom i from atom j, ϕ_{ij} is a pair interaction function and r_{ij} is the distance between atoms i and j.

For homogeneous deformations, integrating the CB strain energy in Eq. (8-14) over the representative volume Ω_0 gives the same result as the energy of an atomic unit cell in a homogeneously deforming crystal. This energetic equivalence forms the basis of the traditional CB hypothesis, in which lattice defects are not allowed; other works, notably the quasicontinuum method [61], have been developed to relieve this restriction. Once the strain energy density is known, continuum stress measures such as the second Piola-Kirchoff stress $\mathbf{S}$, which can be interpreted as the actual force mapped to the undeformed configuration divided by the undeformed area [81], can be defined as

$$\mathbf{S} = 2\frac{\partial \Phi(\mathbf{C})}{\partial \mathbf{C}} = \frac{1}{\Omega_0}\frac{\partial U(\mathbf{C})}{\partial \mathbf{r}}\frac{\partial \mathbf{r}}{\partial \mathbf{C}}, \tag{8-15}$$

while the material tangent modulus C is defined to be

$$\mathcal{C} = 2\frac{\partial \mathbf{S}}{\partial \mathbf{C}}. \tag{8-16}$$

As previously discussed, another key restriction of the CB hypothesis is that all points are assumed to lie in the bulk as $\Phi(\mathbf{C})$ does not account for surface effects. Therefore, the issue at hand is to develop an expression for the energy density along the surfaces of a body, where the potential energy of atoms differs from the bulk due to undercoordination; here, undercoordination is used to describe the fact that atoms at the surfaces of a material have fewer bonding neighbors than atoms that lie within the bulk portion of the material.

As discussed in the introduction, most surface elastic models decompose the total energy of the continuum body into surface and bulk contributions. The uniqueness of the present approach is the usage of the CB approximation in constructing the

surface energy density; we now discuss how the CB approximation can be utilized to approximate the surface energy density.

Equation (8-10) thus represents the decomposition of the total energy of a continuous body into bulk and surface components. For FCC metals, we consider Figure 8-3, which illustrates the bulk/surface decomposition for a ⟨100⟩ oriented FCC crystal with {100} transverse surfaces. For this crystal structure interacting via the EAM potential, there are four layers of surface atoms whose forces deviate from that of an ideal bulk atom. Because the EAM potentials generally employ a fourth shell neighbor cutoff [82], it would appear that the atoms in layers Γ_0^3 and Γ_0^4 feel the same force as those in the bulk. However, the forces on atoms in surface layer Γ_0^3 depend on the electron density of atoms in layers Γ_0^1 and Γ_0^2, while the forces on atoms in surface layer Γ_0^4 depend on the electron density of atoms in layer Γ_0^2. Because atoms in surface layers Γ_0^1 and Γ_0^2 do not have a full complement of neighbors, their electron densities will deviate from the bulk, causing non-bulk forces on atoms in layers Γ_0^3 and Γ_0^4. We concentrate on the effects of forces, and not energies, as the stresses that are needed for the FE formulation in the next section are found by performing a chain rule on the forces, as seen in (8-5).

We first note that for ⟨100⟩ FCC crystals whose interactions are governed by EAM potentials, there exist four non-bulk layers of atoms at the surfaces, as illustrated in Figure 8-3. Thus, we rewrite Eq. (8-10) taking into account the four non-bulk layers to read

$$\sum_{i}^{\text{natoms}} U_i(r) = \int_{\Omega_0^{\text{bulk}}} \Phi(\mathbf{C})\mathrm{d}\Omega + \int_{\Gamma_0^1} \gamma_{\Gamma_0^1}(\mathbf{C})\mathrm{d}\Gamma + \int_{\Gamma_0^2} \gamma_{\Gamma_0^2}(\mathbf{C})\mathrm{d}\Gamma + \int_{\Gamma_0^3} \gamma_{\Gamma_0^3}(\mathbf{C})\mathrm{d}\Gamma + \int_{\Gamma_0^4} \gamma_{\Gamma_0^4}(\mathbf{C})\mathrm{d}\Gamma. \tag{8-17}$$

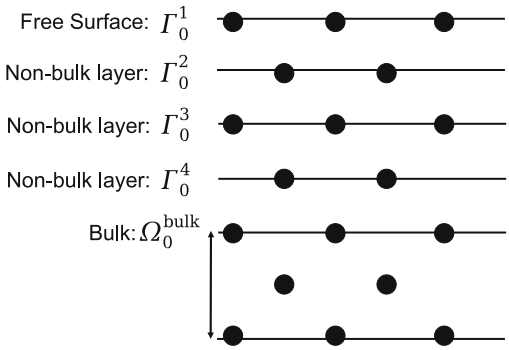


Figure 8-3. Illustration of bulk and non-bulk layers of atoms in a 3D FCC crystal interacting by an EAM potential

Having defined the energy equivalence including both bulk and surface effects, we now determine the surface energy densities $\gamma(\mathbf{C})$. Analogous to the bulk energy density, the surface energy densities $\gamma(\mathbf{C})$ will describe the energy per representative undeformed *area* of atoms at or near the surface of a homogeneously deforming crystal. For FCC metals, choosing a surface unit cell that contains only one atom is sufficient to reproduce the structure of each surface layer. The surface unit cell possesses translational symmetry only in the plane of the surface, unlike the bulk unit cell which possesses translational symmetry in all directions. Thus, the surface energy density for a representative atom in a given surface layer in Figure 8-3 can be written generally as

$$\begin{aligned} \gamma_{\Gamma_0^a}(\mathbf{C}) &= \frac{1}{\Gamma_0}\left(F_i(\bar{\rho}_i) + \phi_i\right), \\ \phi_i &= \frac{1}{2}\sum_{j\neq i}^{n\mathrm{b}_a} \phi_{ij}(r_{ij}(\mathbf{C})), \\ \bar{\rho}_i &= \sum_{j\neq i}^{n\mathrm{b}_a} \rho_j(r_{ij}(\mathbf{C})), \end{aligned} \tag{8-18}$$

where $n\mathrm{b}_a$ are the number of bonds for an atom in surface layer a, and Γ_0 is the representative unit area occupied by a non-bulk atom lying at or near the free surface.

The surface energy densities differ in two ways from the bulk energy density. First, they are normalized by an *area*, Γ_0^a, instead of by a volume. This is necessary to give the correct units of energy when integrating the energy density in (8-17). The second is that an atom in each different surface layer has a different number of bonding neighbors; we again note that while atoms in layers Γ_0^3 and Γ_0^4 have the same number of bonding neighbors as bulk atoms, because the electron densities of the undercoordinated atoms in layers Γ_0^1 and Γ_0^2 are necessary for the force and stress calculations, the atoms in layers Γ_0^3 and Γ_0^4 are treated as non-bulk. For the $\langle 100\rangle$ FCC crystal with $\{100\}$ transverse surfaces considered in this work, atoms in layer Γ_0^1 have 33 neighbors, atoms in Γ_0^2 have 45 neighbors, while atoms in layers Γ_0^3 and Γ_0^4 both have the bulk complement of 54 neighbors.

We note in closing this section that because we have assumed that the energetics of each surface layer can be described by a single representative atom, we have ignored the effects of edge and corner atoms. While these atoms are expected to play an important role in truly small nanostructures [83], the system size at which these effects become significant can easily be described using direct molecular calculations. As will be demonstrated in the numerical examples, the current methodology is geared for larger problems where such edge and corner effects are relatively insignificant, and simultaneously where the system size for fully atomistic calculations becomes prohibitive.

8.2.4. Formulation for Diamond Cubic Lattices

Before presenting the surface Cauchy-Born formulation for diamond cubic lattices, we first present the bulk Cauchy-Born (BCB) model for silicon. The BCB formulation in this work for silicon closely mirrors that of Tang et al. [60], Park and Klein [79] and Park [84]. Because the SCB model for silicon is much easier to understand once the bulk formulation is presented, we present an abbreviated version of the BCB formulation below.

8.2.4.1. Bulk Cauchy-Born Model for Silicon

In the present work, we utilize the T3 form of the Tersoff potential [85] and the resulting parameters. The T3 is named as such due to the fact that two earlier versions of the Tersoff potential suffered from various shortcomings, including not predicting diamond as the ground state of silicon, inaccuracies in the bulk elastic constants, and inaccurate modeling of the {100} surfaces of silicon [86]. The T3 potential energy U can be written as

$$\begin{aligned} U &= \frac{1}{2}\sum_{i\neq j} V_{ij}, \\ V_{ij} &= f_{\mathrm{C}}(r_{ij})\left(f_{\mathrm{R}}(r_{ij}) + b_{ij} f_{\mathrm{A}}(r_{ij})\right), \end{aligned} \tag{8-19}$$

where r_{ij} is the distance between atoms i and j, f_{C} is a cut-off function, which is used to ensure that the Tersoff potential is effectively a nearest neighbor potential, f_{R} is a repulsive function, f_{A} is an attractive function, and b_{ij} is the bond order function, which is used to modify the bond strength depending on the surrounding environment.

The various functions all have analytic forms, which are given as

$$f_{\mathrm{R}}(r_{ij}) = Ae^{-\lambda r ij}, \tag{8-20}$$

$$f_{\mathrm{A}}(r_{ij}) = -Be^{-\mu r ij}, \tag{8-21}$$

$$b_{ij}(1 + \beta^n \zeta_{ij}^n)^{-1/2n}, \tag{8-22}$$

where

$$\zeta_{ij} = \sum_{k\neq i,j} f_{\mathrm{C}}(r_{ik}) g(\theta_{ijk}), \tag{8-23}$$

and

$$g(\theta_{ijk}) = 1 + \frac{c^2}{d^2} - \frac{c^2}{d^2 + (h - \cos\theta_{ijk})^2}, \tag{8-24}$$

where θ_{ijk} represents the angle between a triplet of atoms $i - j - k$.

In order to turn the atomistic potential energy into a form suitable for the BCB approximation, two steps are taken. First, the potential energy is converted into a

strain energy density through normalization by a representative atomic volume Ω_0; Ω_0 can be calculated for diamond cubic (DC) lattices such as silicon by noting that there are 8 atoms in a DC unit cell of volume a_0^3, where $a_0 = 5.432$Å is the silicon lattice parameter. Thus, $\Omega_0 = 8/a_0^3$ for a $\langle 100 \rangle$ oriented silicon crystal. Second, the neighborhood surrounding each atom is constrained to deform homogeneously via continuum mechanics quantities such as the deformation gradient $\mathbf{F}$, or the stretch tensor $\mathbf{C} = \mathbf{F}^T\mathbf{F}$. It is critical to note that due to the usage of nonlinear kinematics through $\mathbf{F}$ and $\mathbf{C}$, the BCB model is a finite deformation, nonlinearly elastic constitutive model that explicitly represents the stretching and rotation of bonds undergoing large deformation.

Silicon is well-known to occur naturally in the DC lattice structure, which is formed through two interpenetrating FCC lattices, where the two FCC lattices are offset by a factor of ($a_0/4$, $a_0/4$, $a_0/4$). The DC lattice is shown in Figure 8-4 which illustrates the interpenetrating FCC lattices. The complication in modeling DC lattices, which will be resolved below, is that the interpenetrating FCC lattices must be allowed to translate with respect to each other. This key restriction can be accommodated through a five-atom unit cell, i.e. atom A and its four neighbors in Figure 8-4b for which the corresponding Tersoff strain energy density Φ. can be written as:

$$\Phi(r_{1j}(\mathbf{C})) = \frac{1}{2\Omega_0} \sum_{j=2}^{5} V_{1j}(r_{1j}(\mathbf{C})), \tag{8-25}$$

where $i = 1$ in (8-25) because atom i is considered the center of the unit cell (see Figure 8-4), and the summation goes over the four nearest neighbor bonds $j = 2, 3, 4, 5$. The full expression for the strain energy density $\Phi(r_{1j})$ can be written as

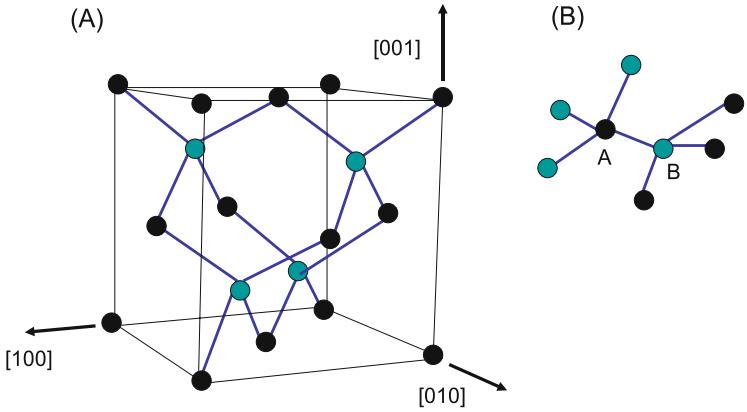


Figure 8-4. Illustration of the diamond cubic lattice structure of silicon. *Black* atoms represent standard FCC unit cell atoms, while *green* atoms represent the interpenetrating FCC lattice. The drawn bonds connect atoms in FCC lattice B to atoms in FCC lattice A

$$\Phi(r_{1j}(\mathbf{C})) = \frac{V_{1j}}{2\Omega_0} = Ae^{-\lambda r_{1j}(\mathbf{C})} - Be^{-\mu r_{1j}(\mathbf{C})}\left(1 + \beta^n\left(\sum_{k\neq i,j} g(\theta_{1jk})\right)^n\right)^{-1/2n} \tag{8-26}$$

where again the multibody effects of the bonding environment are captured through the $g(\theta_{1jk})$ term. We enable the interpenetrating FCC lattices to translate with respect to each other by introducing an internal degree of freedom ξ associated with all neighboring atoms of atom A in Figure 8-4b through the modified bond lengths r_{1j} as

$$r - 1j = |\mathbf{r}_{1j}| = |\mathbf{F}(\mathbf{R}_{1j} + \xi)|, j = 2,3,4,5 \tag{8-27}$$

where $\mathbf{r}_{1j}$ is the deformed bond vector, $\mathbf{R}_{1j}$ is the undeformed bond vector between atoms *1* and *j* and ξ is the shift introduced between the two interpenetrating FCC lattices (i.e. lattices A and B in Figure 8-4) in the undeformed configuration.

The incorporation of the internal degrees of freedom and writing the bond lengths in terms of $\mathbf{F}$ results in a modified strain energy density function as

$$\Phi(\mathbf{C}) = \tilde{\Phi}(\mathbf{C}) = \tilde{\Phi}(\mathbf{C}, \xi(\mathbf{C})). \tag{8-28}$$

Using standard continuum mechanics relationships, we can calculate the second Piola-Kirchoff stress (PK2) as

$$\frac{1}{2}\mathbf{S} = \frac{\partial \Phi}{\partial \mathbf{C}} = \frac{\partial \tilde{\Phi}}{\partial \mathbf{C}} + \frac{\partial \tilde{\Phi}}{\partial \xi}\frac{\partial \xi}{\partial \mathbf{C}}. \tag{8-29}$$

To keep the crystal at an energy minimum, the internal degrees of freedom are constrained to deform according to ξ^*, which leads to the following relationship

$$\frac{\partial \tilde{\Phi}}{\partial \xi^*} = 0, \tag{8-30}$$

and changes the final expression for the PK2 stress to

$$\mathbf{S} = 2\frac{\partial \tilde{\Phi}}{\partial \mathbf{C}}. \tag{8-31}$$

The spatial tangent modulus can be similarly calculated using standard continuum mechanics relations, and can be written as

$$\mathcal{C}_{\mathrm{IJKL}} = M_{\mathrm{IJKL}} - A_{\mathrm{IJ}p}A_{\mathrm{KL}q}(\mathbf{D}^{-1})_{pq}, \tag{8-32}$$

where

$$M_{\mathrm{IJKL}} = 4\frac{\partial^2 \tilde{\Phi}}{\partial C_{\mathrm{IJ}} \partial C_{\mathrm{KL}}}, \tag{8-33}$$

$$D_{pq} = \frac{\partial^2 \tilde{\Phi}}{\partial \xi_p^* \partial \xi_q^*},$$

$$A_{\mathrm{IJ}p} = 2\frac{\partial^2 \tilde{\Phi}}{\partial C_{\mathrm{IJ}} \partial \xi_p^*}.$$

8.2.4.2. Surface Cauchy-Born Model for Silicon

In this section, we present the formulation by which surface stresses are accounted for through an extension of the BCB model we call the surface Cauchy-Born (SCB) model. The SCB model was developed previously for both FCC crystals [87,88] and for DC lattices [79]. We therefore briefly summarize the relevant aspects of the SCB model for silicon [79] in this section. We first note that the total energy of a nanostructure can be written as the sum of bulk and surface terms

$$\sum_{\alpha=1}^{n_{\mathrm{atoms}}} U_\alpha(r) \approx \int_{\Omega_0^{\mathrm{bulk}}} \Phi(\mathbf{C}) \mathrm{d}\Omega + \int_{\Gamma_0} \gamma(\mathbf{C}) \mathrm{d}\Gamma, \tag{8-34}$$

where $U_\alpha(r)$ represents the potential energy for each atom α, $\Phi(\mathbf{C})$ is the bulk energy density previously defined in (8-26) and $\gamma(\mathbf{C})$ is the surface energy density. The issue then is to determine a representation for the surface unit cell that will be used to calculate the surface energy density $\gamma(\mathbf{C})$.

We accomplish this through the nine atom surface unit cell for unreconstructed {100} silicon surfaces shown in Figure 8-5. The rationale for this particular unit cell

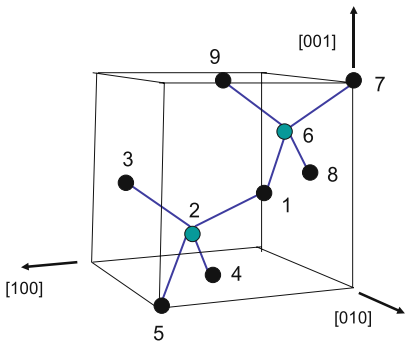


Figure 8-5. Illustration of the nine atom surface unit cell for the surface with a [010] normal of a diamond cubic crystal. *Black* atoms represent FCC lattice A, while *green* atoms represent the interpenetrating FCC lattice B. The drawn bonds connect atoms in FCC lattice B to atoms in FCC lattice A

arises because atoms 2 and 6 both have a full complement of neighbors, and thus represent a distinct FCC lattice B. The atoms neighboring atoms 2 and 6 therefore must be part of the interpenetrating FCC lattice A, and thus should be able to translate with respect to atoms 2 and 6. Therefore, we assign an internal degree of freedom ξ^s, where the superscript s designates an internal surface degree of freedom, to all the black atoms (1, 3, 4, 7, 8, 9) of FCC lattice A in Figure 8-5.

The resulting strain energy density γ for the surface unit cell seen in Figure 8-5 can thus be written as

$$\gamma = \frac{1}{\Gamma_0}\left(\sum_{j=2,\,6} V_{1j}(r_1 j) + \sum_{k=1,\,7,\,8,\,9} V_{6k}(r_{6k}) + \sum_{m=1,\,3,\,4,\,5} V_{2m}(r_{2m})\right) \tag{8-35}$$

where Γ_0 is the area per atom on the surface. Following (8-27), we express the bond lengths for the surface unit cell as

$$\begin{aligned} r_{1j} &= |\mathbf{r}_{1j}| = |\mathbf{F}(\mathbf{R}_{1j} + \xi^s)|, j = 2, 6 \\ r_{6k} &= |\mathbf{r}_{6k}| = |\mathbf{F}(\mathbf{R}_{6k} + \xi^s)|, k = 1, 7, 8, 9 \\ r_{2m} &= |\mathbf{r}_{2m}| = |\mathbf{F}(\mathbf{R}_{2m} + \xi^s)|, m = 1, 3, 4, 5 \end{aligned} \tag{8-36}$$

Incorporating the bond lengths that have been modified by the deformation gradient $\mathbf{F}$ and the internal degrees of freedom ξ^s in (8-36) creates a modified surface energy density $\tilde{\gamma}(\mathbf{C})$ from (8-35), where the surface energy density can be modified analogously to the procedure outlined previously for the bulk energy density in (8-29) and (8-30) to enforce the energy minimizing condition

$$\frac{\partial \tilde{\gamma}}{\partial \tilde{\xi}^s} = 0 \tag{8-37}$$

where $\tilde{\xi}^s$, similar to the meaning in the bulk case in (8-30), represents the deformation of the surface internal degrees of freedom necessary to minimize the surface energy. Using the modified surface energy density $\tilde{\gamma}(\mathbf{C})$, we arrive at the expression for the surface PK2 stress $\mathbf{S}^s(\mathbf{C})$, where the superscript s here and below indicates surface values

$$\mathbf{S}^s(\mathbf{C}) = 2\frac{\partial \tilde{\gamma}(\mathbf{C})}{\partial \mathbf{C}}. \tag{8-38}$$

Similarly, the surface tangent modulus can be written as

$$\mathcal{C}^s_{\mathrm{IJKL}} = \mathbf{M}^s_{\mathrm{IJKL}} - A^s_{\mathrm{IJ}p} A^s_{\mathrm{KL}q} (\mathbf{D}^{-1})^s_{pq}, \tag{8-39}$$

where

$$M^s_{\mathrm{IJKL}} = 4\frac{\partial^2 \tilde{\gamma}}{\partial C^s_{\mathrm{IJ}} \partial C^s_{\mathrm{KL}}}, \tag{8-40}$$

$$D^s_{pq} = \frac{\partial^2 \tilde{\gamma}}{\partial \tilde{\xi}^s_p \partial \tilde{\xi}^s_q},$$

$$A^s_{IJp} = 2\frac{\partial^2 \tilde{\gamma}}{\partial C^s_{IJ} \partial \tilde{\xi}^s_p}.$$

8.3. FINITE ELEMENT FORMULATION AND IMPLEMENTATION

8.3.1. Variational Formulation

Having defined the surface energy densities $\gamma(\mathbf{C})$ for each non-bulk layer of atoms near the surface, we can immediately write the total potential energy Π. of the system including external loads $\mathbf{T}$ as

$$\Pi(\mathbf{u}) = \int_{\Omega_0^{\mathrm{bulk}}} \Phi(\mathbf{C}) d\Omega + \int_{\Gamma_0} \gamma_{\Gamma_0}(\mathbf{C}) \mathrm{d}\Gamma - \int_{\Gamma_0} (\mathbf{T} \cdot \mathbf{u}) \mathrm{d}\Gamma. \tag{8-41}$$

We note that the surface integral involving γ in (8-41) is a general expression for the surface energy density; note that the surface energy density could involve separate terms for each surface layer, as is the case for FCC metals as in (8-17). In order to obtain a form suitable for FE calculations, we introduce the standard discretization of the displacement field $\mathbf{u}(\mathbf{X})$ using FE shape functions as

$$\mathbf{u}(\mathbf{X}) = \sum_{I=1}^{nn} N_1(\mathbf{X}) \mathbf{u}_I, \tag{8-42}$$

where N_I are the shape or interpolation functions, nn are the total number of nodes in the discretized continuum, and $\mathbf{u}_I$ are the displacements of node I [81,89]. Substituting Eqs. (8-14) and (8-18) into Eq. (8-41) and differentiating gives the minimizer of the potential energy and also the FE nodal force balance [81]

$$\frac{\partial \Pi}{\partial \mathbf{u}_I} = \int_{\Omega_0^{\mathrm{bulk}}} \mathbf{B}^T \mathbf{S} \mathbf{F}^T d\Omega + \int_{\Gamma_0} \mathbf{B}^T \tilde{\mathbf{S}} \mathbf{F}^T d\Gamma - \int_{\Gamma_0} N_I \mathbf{T} d\Gamma, \tag{8-43}$$

where $\mathbf{S}$ is the second Piola-Kirchoff stress due to the bulk strain energy and $\mathbf{B}^T = \left(\frac{\partial N_I}{\partial \mathbf{X}}\right)^T$; $\tilde{\mathbf{S}}$ can be loosely labeled as surface Piola-Kirchoff stresses that can be found using Eqs. (8-18) and (8-5) to be of the form

$$\tilde{\mathbf{S}}(\mathbf{C}) = 2\frac{\partial \gamma \Gamma_0(\mathbf{C})}{\partial \mathbf{C}} = \frac{1}{\Gamma_0} \frac{\partial U(\mathbf{C})}{\partial \mathbf{r}} \frac{\partial \mathbf{r}}{\partial \mathbf{C}}. \tag{8-44}$$

The surface Piola-Kirchoff stresses differ from those in the bulk because the normalization factor is an area, instead of a volume. In addition, the surface Piola-Kirchoff stresses $\tilde{\mathbf{S}}(\mathbf{C})$ are 3×3 tensors with normal components which allow surface relaxation due to undercoordinated atoms lying at material surfaces; this result differs from the traditional definition of surface stress [33] which is a 2×2 tensor with only tangential components.

The normal components arise in the present approach because the atoms that constitute the surface unit cells lack proper atomic coordination in the direction normal to the surface; therefore, the atomistic forces that are normalized to stresses in Eq. (8-44) are also out of balance in the normal direction. Thus, surface relaxation is necessary in the normal direction to regain an equilibrated state. Further details on the numerical implementation of the surface layers can be found in [87].

In contrast, traditional surface elastic approaches [41, 55] utilize a 2×2 stress tensor which depends only on the tangential components of the deformation, and thus cannot explain surface relaxation effects. Solving (8-43) requires an iterative process to solve the nonlinear system of equations. The purpose of the iterative procedure is to determine the unknown FE nodal displacements $\mathbf{u}_I$ that minimize the energy functional $\Pi(\mathbf{u})$.

What has been accomplished in (8-43) is a systematic manner of obtaining continuum stress measures by calculating the system potential energy as a function of bulk and surface components. By correctly calculating the system energy, standard continuum mechanics relationships can be utilized to derive stress measures for usage in FE computations. The salient feature of equation (8-43) is that as the surface area to volume ratio becomes larger, the surface area terms will dominate the energetic expression. Because the stresses required for the FE internal forces are calculated by differentiating the strain energy density, correctly accounting for the surface energy will naturally lead to the correct forces on surface nodes.

This situation is exactly that which occurs in nanoscale materials such as nanowires, quantum dots, nanoparticles and nanobeams; for these small scale structures, the finite surface energies will create surface stresses that can cause both surface relaxation into the bulk, as well as unique mechanical properties caused by the need to overcome the intrinsic surface stresses [90–92]. In contrast, if the volume of the material is significantly larger than the surface area, then the potential energy from the surface terms will be insignificant compared to the volumetric potential energy, and the material will feel no effect from the surface stresses. Thus, this model degenerates to a bulk Cauchy-Born model as the length scale of the material increases.

8.3.2. Finite Element Eigenvalue Problem for Nanowire Resonant Frequencies

In order to calculate the resonant frequencies of the nanowires, it will be necessary to solve a standard finite element eigenvalue problem. The eigenvalue problem is obtained from the equation describing the eigenvalue problem for continuum elastodynamics, which is written as

$$(\mathbf{K} - \omega^2 \mathbf{M})\mathbf{u} = 0, \tag{8-45}$$

where **M** is the mass matrix and **K** is the stiffness matrix of the discretized FE equations; the solution of the eigenvalue problem described in Eq. (8-45) gives the resonant frequencies f, where $f = \omega/2\pi$ and the corresponding mode shapes **u**. We note that the stiffness matrix **K** contains the effects of both material and geometric nonlinearities through a consistent linearization about the finitely deformed configuration.

We emphasize that the addition of the surface energy terms in Eq. (8-34) leads naturally to the incorporation of the surface stresses in the FE stiffness matrix **K**, which then leads to the dependence of the resonant frequencies f on the surface stresses. The eigenvalue problem was solved using the Sandia-developed package Trilinos [93], which was incorporated into the simulation code Tahoe [94].

8.4. APPLICATIONS OF SURFACE CAUCHY-BORN MODEL

We proceed now to demonstrate the capabilities of the proposed SCB model. We first validate the model through comparison with a full scale atomistic calculation for FCC metal nanowires. We then move forward to demonstrate what can be learned about silicon nanowires using the SCB model, particularly about surface stress effects on their resonant frequencies, which will be critical for the design and analysis of NEMS.

8.5. DIRECT SURFACE CAUCHY-BORN/MOLECULAR STATICS COMPARISON

The numerical examples for the direct SCB/molecular statics comparison utilizes geometries similar to that shown in Figure 8-6, which illustrates a gold nanowire with square cross section of length $a$and longitudinal length h. All wires had a $\langle 100 \rangle$ longitudinal orientation with $\{100\}$ transverse side surfaces, and were subject to the same boundary conditions; the left $(-x)$ surface of the wires were fixed, while the right $(+x)$ surface of the wires were constrained to move only in the x-direction. All simulations, both molecular statics (MS) for the benchmark atomistics and the FE for the SCB were performed using the stated boundary conditions without additional

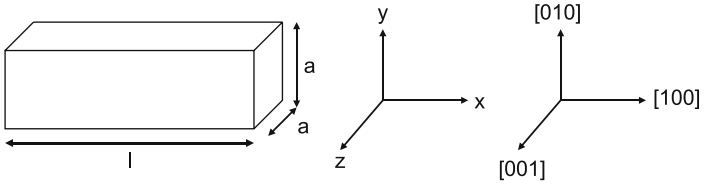


Figure 8-6. Nanowire geometry considered for surface-stress-driven relaxation examples

external loading and without periodic boundary conditions. Therefore, all deformation observed in the examples is caused by the effects of surface stresses. All SCB calculations utilized regular meshes of 8-node hexahedral elements.

The atomistic interactions were based on the EAM, with gold being the material for all problems using the parameters of Foiles [82], while the same parameters were used to calculate the bulk and surface stresses needed for the SCB simulations; single crystal gold nanowires were considered in all cases. Care was taken to consider nanowires with sizes large enough such that surface-stress-driven phase transformations [34] or reorientations [95], which have been predicted in gold nanowires with cross sections smaller than about 2 nm, did not occur. All simulations, for both FE and MS, were performed quasistatically to find energy minimizing positions of either the atoms or the FE nodes accounting for the surface stresses.

The first example illustrates a direct comparison between a benchmark MS calculation and a SCB calculation. For this, the gold nanowire was comprised of 145261 atoms with dimensions of $24.48 \times 9.792 \times 9.792$ nm. The equivalent SCB model contained a regular mesh of 576 finite elements and 833 nodes, leading in a 99.4% reduction in the number of degrees of freedom; because similar mesh densities were used for all simulations shown in this work, similar reductions in the required degrees of freedom for the SCB calculations were achieved in all cases.

Due to the surface stresses, the $+x$ edge of the wire contracts upon relaxation, resulting in an overall state of compressive strain in the wire. For both the MS and SCB calculations, the compressive strain was calculated by measuring the displacement at the center of the $+x$ surface at $(+x, 0, 0)$. This was done because, as seen in Figure 8-7, the corners of the $+x$ surface of the nanowire have a greater contraction than the center of the $+x$ surface because they have the greatest degree of undercoordination. The SCB calculation predicted a compressive strain to relaxation of -0.91%, while the MS calculation predicted a contraction of -0.83%.

The fact that the SCB can predict the compressive relaxation is strengthened by comparative calculations for the relaxed and unrelaxed surface energy for both the SCB and MS systems for the same EAM potential; the unrelaxed surface energies γ_{ur} were found to be 0.975 J/m^2 for the MS system, and 0.973 J/m^2 for the SCB system for the {100} surface of gold using Foiles et al. [82] potential. The relaxed surface energies γ_r were found to be 0.914 J/m^2 for the MS system, and 0.932 J/m^2 for the SCB system; the slight overestimation of the relaxed surface energy by the SCB correlates correctly with the higher relaxation strain in the above numerical example.

The overall contours of the x and y-displacements are shown in Figures 8-7 and 8-8. As can be seen, the SCB calculations reproduce well the displacement fields in both the x and y-directions, including the compressive relaxation in the x-direction at the $+x$ edge of the nanowire, which then causes expansion of the nanowire in the y and z-directions. The y-displacement was calculated at the center of the $+y$ surface at $(0, +y, 0)$ and was compared for both the SCB and MS calculations; the SCB predicted an expansion of 0.18% while the MS calculation predicted an expansion of 0.17%.

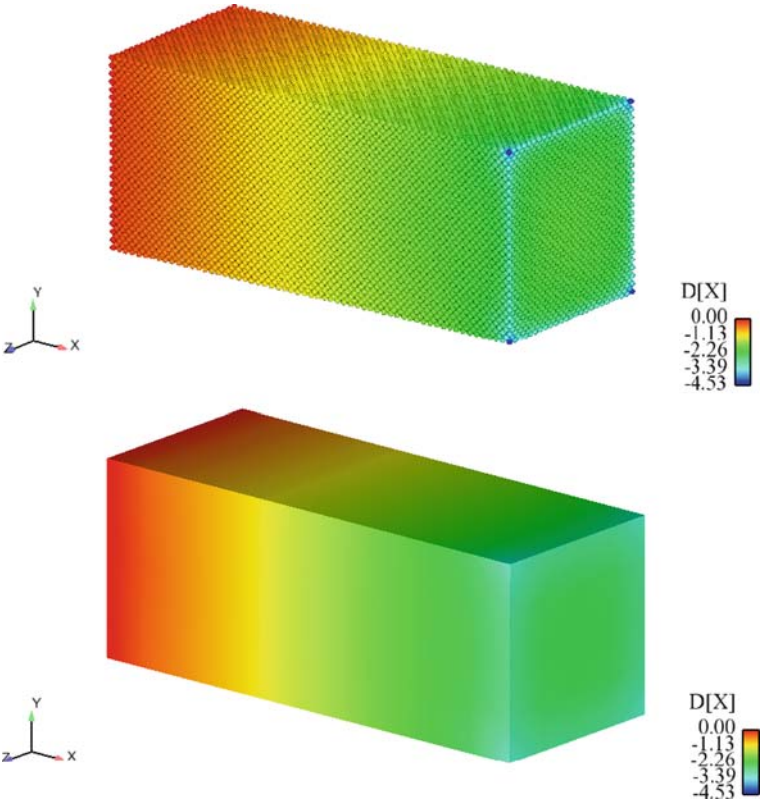


Figure 8-7. Comparison of x-displacements for $24.48 \times 9.792 \times 9.792$ nm gold nanowire for (*top*) MS and (*bottom*) SCB calculations

The snapshots of the x and y-displacements shown in Figures 8-7 and 8-8 serve to highlight both the strengths and the weaknesses of the current version of the SCB method. As mentioned above, the SCB method clearly captures in a qualitative sense the overall relaxed configuration for the gold nanowire, at a greatly reduced computational cost as compared to the MS simulation. On the other hand, the results of the MS simulations show that the corners and edges of the nanowire experience a considerably different deformation than the surfaces and the bulk. Because the SCB method as currently formulated does not account for corner and edge effects, the deformation of those areas is captured in an average sense due to the deformation of the adjacent surfaces.

8.6. SURFACE STRESS EFFECTS ON THE RESONANT PROPERTIES OF SILICON NANOWIRES

We now demonstrate the applicability and utility of the SCB model in predicting surface stress effects on the resonant properties of silicon nanowires [84]. All numerical

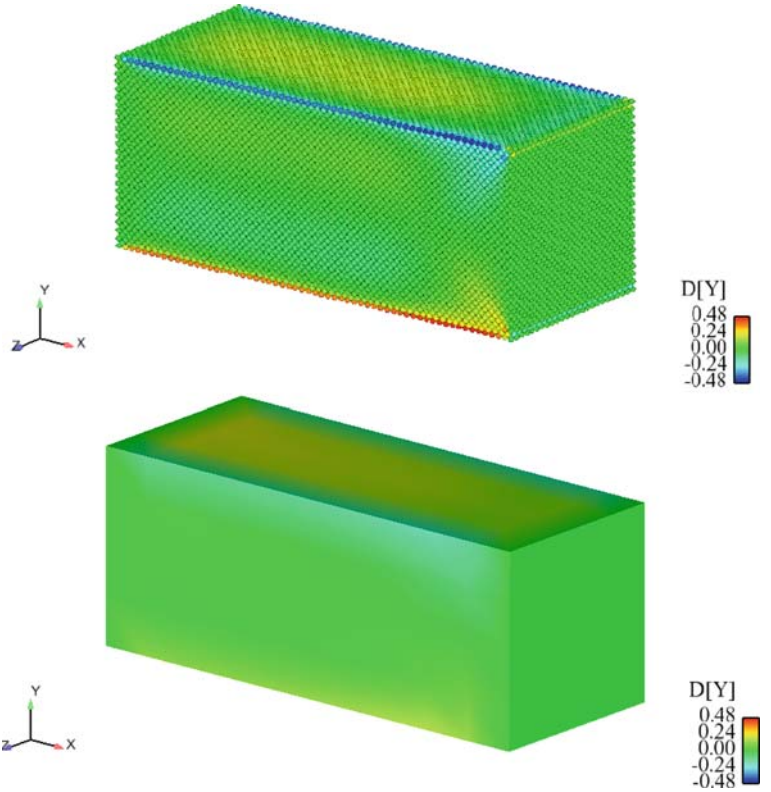


Figure 8-8. Comparison of *y*-displacements for 24.48 × 9.792 × 9.792 nm gold nanowire for (*top*) MS and (*bottom*) SCB calculations

examples were performed on three-dimensional, single crystal silicon nanowires of length *l* that have a square cross section of width *a* as illustrated in Figure 8-6. Three different parametric studies are conducted in this work, which consider nanowires with constant cross sectional area (CSA), constant length and constant SAV (SAV); the geometries are summarized in Table 8-1.

Table 8-1. Summary of geometries considered: constant SAV ratio (SAV), constant length, and constant cross sectional area (CSA). All dimensions are in nm

Constant SAV	Constant Length	Constant CSA
64 × 16 × 16	240 × 8 × 8	64 × 16 × 16
110 × 15.2 × 15.2	240 × 12 × 12	128 × 16 × 16
170 × 14.9 × 14.9	240 × 18 × 18	256 × 16 × 16
230 × 14.7 × 14.7	240 × 24 × 24	384 × 16 × 16
290 × 14.5 × 14.5	240 × 30 × 30	512 × 16 × 16

All wires had a ⟨100⟩ longitudinal orientation with unreconstructed {100} transverse surfaces, and had either fixed/free (cantilevered) boundary conditions, where the left ($-x$) surface of the wire was fixed while the right ($+x$) surface of the wire was free, or fixed/fixed boundary conditions, where both the left ($-x$) and right ($+x$) surfaces of the wire were fixed. All FE simulations were performed using the stated boundary conditions without external loading, and utilized regular meshes of 8-node hexahedral elements. The bulk and surface energy densities in Eqs. (8-26) and (8-35) were calculated using Tersoff T3 parameters [85], while the bulk and surface FE stresses were found using Eqs. (8-31) and (8-38).

Regardless of boundary condition, the nanowires are initially out of equilibrium due to the presence of the surface stresses. For fixed/free nanowires, the free end expands in tension to find an energy minimizing configuration under the influence

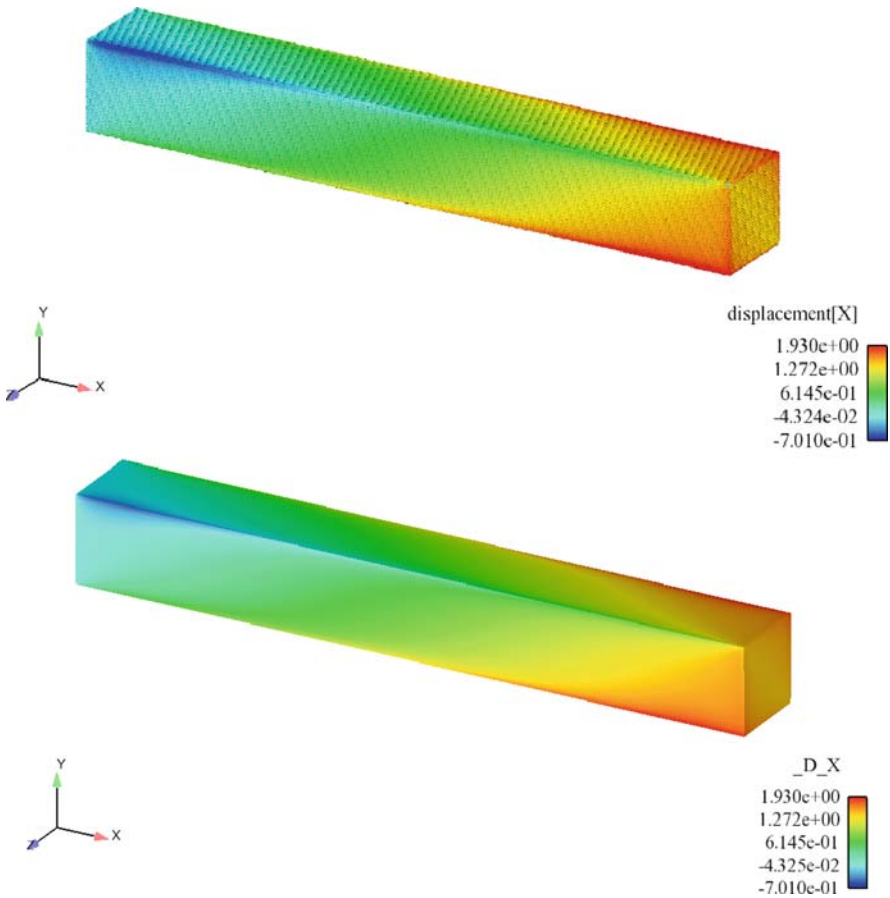


Figure 8-9. (*Color online*) Minimum energy configuration of a fixed/free 128 × 16 × 16 nm silicon nanowire due to surface stresses as predicted by (*top*) MS calculation, (*bottom*) SCB calculation

of surface stresses. To illustrate this, we compare the energy minimized positions of the 128 × 16 × 16 nm fixed/free nanowire using the SCB model to a benchmark molecular statics (MS) calculation performed using the Tersoff T3 potential with LAMMPS [96] MS code. As seen in Figure 8-9, the SCB model, which required only 16393 FE nodes, gives a very accurate description of the minimum energy configuration due to surface stresses as compared to the MS calculation, which required more than 1.7 million atoms. We note that the tensile strain induced in the nanowires due to the surface stresses is about 0.1%.

As noted previously, no external forces were applied to obtain the results seen in Figure 8-9; all deformation is solely due to surface stresses. In analyzing the results in Figure 8-9, we emphasize that the SCB model accurately predicts the tensile expansion of the free end due to surface stresses, in addition to capturing the inhomogeneous nature of the tensile expansion, which occurs due to the non-centrosymmetric nature of the DC silicon lattice. The results are in agreement with first principles calculations [86,97], which also indicate that ⟨100⟩ silicon nanowires with {100} surfaces have compressive surface stresses that cause the nanowires to expand.

Fixed/fixed nanowires, on the other hand, are constrained such that the nanowire is unable to expand due to the boundary conditions. The boundary condition constraint therefore causes the minimum energy configuration of fixed/fixed nanowires to be a state of compression, which we will demonstrate is critical to understanding how surface stresses and boundary conditions couple to alter the resonant properties of fixed/fixed nanowires as compared to continuum beam theory predictions.

Once the minimum energy configuration for either boundary condition is known, the eigenvalue problem described in Eq. (8-45) is solved using the FE stiffness matrix from the equilibrated (deformed) nanowire configuration to find the resonant frequencies. Resonant frequencies were also found using the standard BCB model (without surface stresses) on the same geometries for comparison to quantify how surface stresses change the resonant frequencies as compared to the bulk material for a given geometry and boundary condition. For all resonant frequencies reported in this work, the fundamental, or lowest mode frequencies corresponded to a standard bending mode of deformation.

8.6.1. Constant Cross Sectional Area

To validate the accuracy of the calculations for the BCB material, we compare in Tables 8-2 and 8-3 the BCB and SCB resonant frequencies to those obtained using the well-known analytic solutions for the fundamental resonant frequency for both fixed/free (cantilevered) and fixed/fixed beams [98]. For the fixed/free beam:

$$f_0 = \frac{B_0^2}{2\pi l^2}\sqrt{\frac{EI}{\rho A}}, \tag{8-46}$$

where $B_0 = 1.875$ for the fundamental resonant mode, E is the modulus for silicon in the ⟨100⟩ direction, which can be found to be 90 GPa [85], I is the moment of

Table 8-2. Summary of constant CSA nanowire fundamental resonant frequencies for fixed/free boundary conditions as computed from: (1) The analytic solution given by Eq. (8-46), (2) Bulk Cauchy-Born (BCB), and (3) Surface Cauchy-Born (SCB) calculations. All frequencies are in MHz, the nanowire dimensions are in nm

Geometry	Equation 46	BCB	SCB
$64 \times 16 \times 16$	3933	3912	4008
$128 \times 16 \times 16$	983	990	1013
$256 \times 16 \times 16$	246	248	253
$384 \times 16 \times 16$	109	110	112
$512 \times 16 \times 16$	62	62	63

Table 8-3. Summary of constant CSA nanowire fundamental resonant frequencies for fixed/fixed boundary conditions as computed from: (1) The analytic solution given by Eq. (8-47), (2) Bulk Cauchy-Born (BCB), and (3) Surface Cauchy-Born (SCB) calculations. All frequencies are in MHz, the nanowire geometry is in nm.

Geometry	Equation 47	BCB	SCB
$64\times16\times16$	24842	21618	22165
$128\times16\times16$	6211	6074	6166
$256\times16\times16$	1553	1565	1528
$384\times16\times16$	690	698	635
$512\times16\times16$	388	393	317

inertia, l is the nanowire length, A is the cross sectional area and ρ is the density of silicon. The FE calculations used to calculate the BCB and SCB resonant frequencies involved regular meshes of 8-node hexahedral elements; the mesh sizes ranged from about 8000 to 65,000 nodes for the constant CSA nanowires considered.

The BCB resonant frequencies compare quite well to those predicted by the analytic formula, with increasing accuracy for increasing aspect ratio l/a, as would be expected from beam theory. We note that the SCB resonant frequencies are consistently larger than the BCB resonant frequencies and thus the analytic solution; reasons for this trend will be discussed later.

For the fixed/fixed beam, the analytic solution is given as [98]

$$f_0 = \frac{i^2\pi}{2l^2}\sqrt{\frac{EI}{\rho A}}, \tag{8-47}$$

where $i \approx 1.5$ is a mode shape factor for fixed/fixed beams. Table 8-3 shows that the BCB and analytic solutions again agree nicely. However, in contrast to the fixed/free case, the SCB resonant frequencies are found to decrease with increasing aspect ratio relative to the bulk material; again, reasons for this will be discussed later. A key point to emphasize here is that due to the accuracy of the BCB results for both boundary conditions as compared to the analytic solutions, normalizing the SCB resonant frequencies by the BCB resonant frequencies can be considered to be equivalent to normalizing by the solution expected from continuum beam theory.

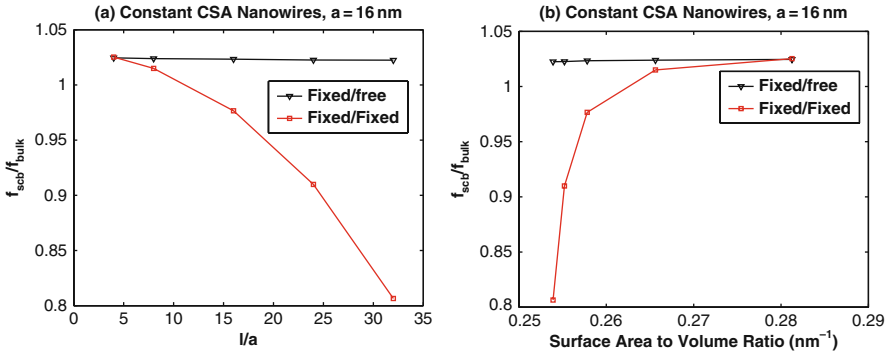


Figure 8-10. Normalized resonant frequencies for constant CSA silicon nanowires

Figure 8-10a shows the normalized resonant frequencies $f_{\text{scb}}/f_{\text{bulk}}$ plotted against both the nanowire aspect ratio l/a and the SAV ratio, for both fixed/fixed and fixed/free boundary conditions. As can be observed, the surface stress effects on the resonant frequencies depend strongly upon the corresponding boundary conditions. For the fixed/free nanowires, the resonant frequencies predicted using the SCB model are about 2% higher than those of the BCB model for all aspect ratios. In contrast, the fixed/fixed nanowires show completely different behavior. In that case, the resonant frequencies predicted by the SCB model dramatically decrease with increasing aspect ratio l/a, with the resonant frequencies due to surface stress decreasing to nearly 20% lower than the corresponding bulk material when the aspect ratio $l/a > 30$.

The resonant frequency calculations are also plotted with respect to the SAV ratio in Figure 8-10b. As can be seen, the fixed/free nanowires show little variation with the SAV ratio, while the fixed/fixed nanowires show a decrease in resonant frequency with decreasing SAV ratio.

8.6.2. Constant Length

We next investigate the resonant frequencies of nanowires in which the length of the nanowire was fixed at 240 nm, while the square cross section was varied in size. The FE calculations to determine the resonant frequencies required mesh sizes ranging from about 13,000 nodes for the smallest (8 nm) cross section to about 71,000 nodes for the largest (30 nm) cross section considered.

As with the constant CSA nanowires, we plot the $f_{\text{scb}}/f_{\text{bcb}}$ ratio against both the aspect ratio l/a and the SAV ratio in Figure 8-11a. When plotted against the aspect ratio l/a, the trends for the constant length nanowires are similar to those of the constant CSA nanowires, particularly for the fixed/fixed boundary conditions, for which surface stresses cause the resonant frequencies to decrease rapidly with increasing l/a. In fact, for $l/a = 30$ for the 8 nm cross section nanowire, surface stresses cause the resonant frequency to be less than 65% of the bulk value. The surface stresses cause a slightly different trend for the fixed/free case. There, the

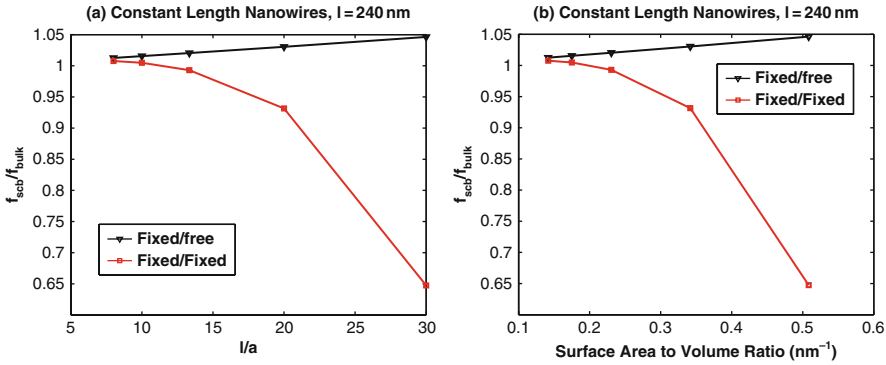


Figure 8-11. Normalized resonant frequencies for constant length silicon nanowires

resonant frequencies are observed to increase slightly with respect to the bulk value with increasing l/a, while the trend was a very minute decrease in the constant CSA case.

However, when plotted against the SAV ratio, as in Figure 8-11b, the results for constant length nanowires differ strongly from the constant CSA nanowires. In particular, the result is most noticeable for the fixed/fixed nanowires; in the constant CSA case, surface stresses caused an *increase* in resonant frequency with increasing SAV ratio. However, for the constant length nanowires, the opposite trend is observed; the surface stresses cause the resonant frequencies to *decrease* with increasing SAV ratio. The trends are also reversed, though not as dramatically, for the fixed/free boundary conditions.

8.6.3. Constant Surface Area to Volume Ratio

Due to the variation in surface stress and boundary condition effects on the nanowire resonant frequencies, we consider those coupled effects for nanowires that have the same SAV ratio, $0.28\,\text{nm}^{-1}$. The FE mesh sizes ranged in this case from about 15,000 nodes (for the 15.2 nm cross section nanowire) to about 41,000 nodes (for the 14.5 nm cross section nanowire).

Because the SAV ratio is kept constant, we plot the resonant frequencies for both boundary conditions only against the nanowire aspect ratio l/a in Figure 8-12. Figure 8-12 thus shows one of the fundamental findings of this work, in that the resonant frequencies of fixed/fixed silicon nanowires do not, due to surface stresses, depend on the SAV ratio. The results for the fixed/free nanowires are more ambiguous judging solely from Figure 8-12. However, Figures 8-10a and 8-11a indicate that the resonant frequencies of fixed/free silicon nanowires, similar to fixed/fixed silicon nanowires, do not scale according to SAV ratio.

In particular, in all cases, it appears that the nanowire aspect ratio l/a is a much stronger predictor of how the boundary conditions and surface stresses couple to vary the resonant frequencies as compared to the corresponding bulk material than

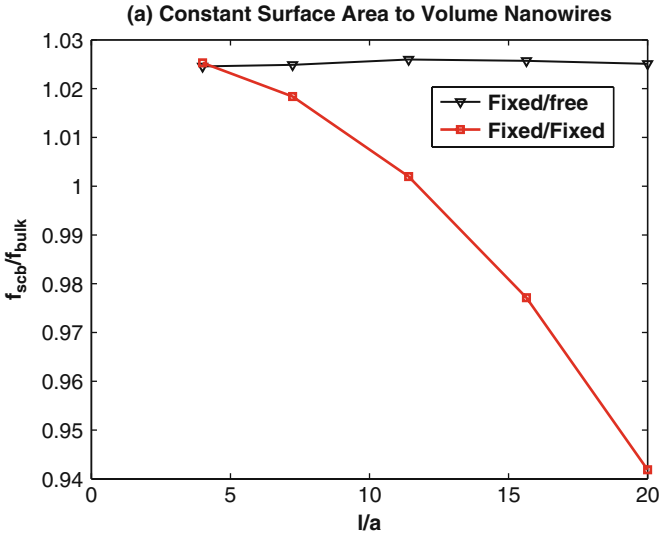


Figure 8-12. Normalized resonant frequencies for constant SAV silicon nanowires

the SAV ratio. This finding corresponds to results recently published by Verbridge et al. [22], and Petrova et al. [10]. The Verbridge work analyzed the resonant properties of SiN nanostrings, with cross sectional dimensions around 100 nm. While the surface stress effects observed in the present work are unlikely to have a significant impact on 100 nm cross section nanowires, it is interesting that even when surface effects become significant, as they do for the nanowires considered in the present work, that the resonant frequencies, and thus the elastic properties, are largely independent of SAV ratio. The Petrova work offers a comparison at a different length scale (cross sections on the order of 10-20 nm) and for a different material, gold. However, that work also found weak dependence of the resonant frequencies and thus elastic properties on the SAV ratio; these results, on different materials at different sizes, lend credibility to the results obtained in the present work.

8.7. DISCUSSION AND ANALYSIS

We now present an analysis of the boundary condition and surface stress effects on the nanowire elastic properties, and in particular the Young's modulus. To calculate the Young's modulus, we utilize the beam theory expressions that relate the resonant frequencies to the modulus in Eqs. (8-46) and (8-47). The beam theory expressions for the modulus are utilized as they are also ubiquitous in the experimental literature to calculate the Young's modulus for nanostructures [6,30,29,22,10,99].

Figure 8-13 depicts the variation in the Young's modulus, as normalized by the bulk value, for both the fixed/fixed and fixed/free constant CSA nanowires. As can be observed, the Young's modulus for the fixed/free case shows about a 5% variation

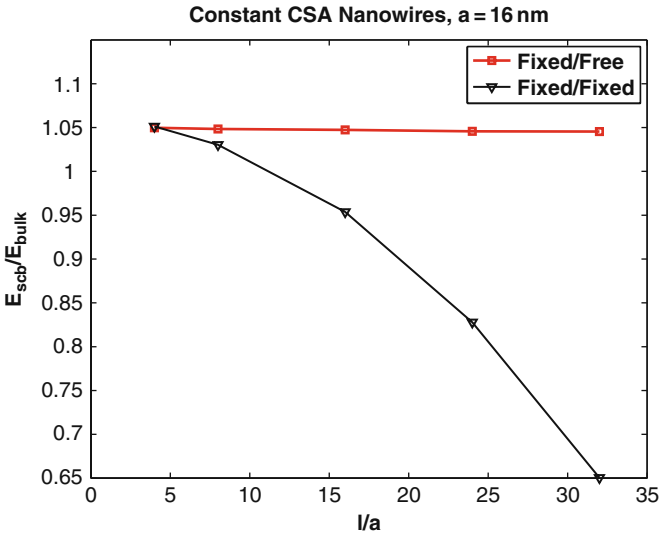


Figure 8-13. Normalized Young's modulus for both fixed/fixed and fixed/free boundary conditions for constant CSA nanowires

from the bulk value, which can be expected from the fact that the fixed/free resonant frequencies, when bulk normalized, also showed a small increase with respect to the bulk resonant frequencies.

However, there is a dramatic variation in the fixed/fixed Young's modulus with increasing aspect ratio l/a, as shown in Figure 8-13. In particular, due to the nature of the resonance formula in (8-47), the modulus that is calculated is actually significantly reduced as compared to the bulk Young's modulus than the bulk-normalized resonant frequencies. For example, $f_{\text{scb}}/f_{\text{bulk}} = 0.81$ for $l/a = 30$, as seen in Figure 8-10a. However, when surface stresses are accounted for, the Young's modulus drops to only 65% of the bulk modulus when $l/a = 32$.

Furthermore, this observed reduction of the Young's modulus has been observed in other theoretical studies for fixed/fixed silicon nanowires. In particular, we note the recent density functional theory studies by Lee and Rudd for ultrasmall (< 4 nm) fixed/fixed silicon nanowires [97], which also predicted a decrease in Young's modulus due to the fact that the surface stresses in conjunction with the fixed/fixed boundary conditions cause the nanowire to exist in a state of compression; we note that the variation of the Young's modulus accounting for length was not performed in that work. Molecular dynamics simulations of the resonant frequencies of fixed edge silicon oxide nanoplates by Broughton et al. [100] also revealed a distinct reduction in the resonant frequencies with decreasing size.

We also seek to quantify the variations due to surface stresses in the resonant frequencies for the fixed/fixed case. To do so, Figure 8-14, which plots the normalized resonant frequencies $f_{\text{scb}}/f_{\text{bulk}}$ for all fixed/fixed nanowires (constant CSA, length, SAV) against the nanowire aspect ratio l/a, demonstrates one of the major findings

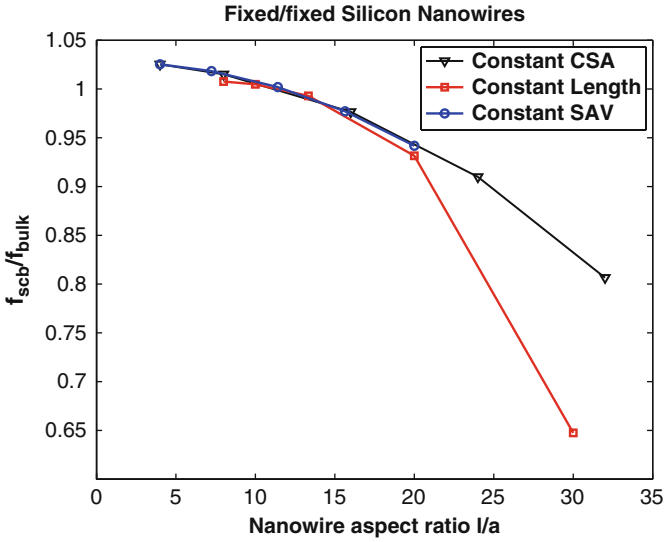


Figure 8-14. Normalized resonant frequencies for fixed/fixed constant CSA, SAV and length nanowires plotted against the nanowire aspect ratio

of this work. As can be seen, for the nanowire sizes considered in this work, the resonant frequencies for all nanowires as compared to the resonant frequencies of the bulk material overlap on a similar curve as a function of the aspect ratio, with the trend being a decreasing resonant frequency with increasing aspect ratio. The enhanced effect of surface stresses for the constant length nanowire with aspect ratio of $l/a = 30$ is likely due to the fact that it was the smallest cross section considered, i.e. 8 nm, where the surface stress effects are particularly strong. Figure 8-14 can therefore serve as a design guide for predicting how surface stresses will change the resonant frequencies of nanowires as compared to the continuum beam theory in Eq. (8-47) which does not account for surface effects.

We attempted to determine similar relationships for the fixed/free nanowires in linking the observed variations of the nanowire resonant frequencies due to surface stresses to geometric parameters. Unfortunately, as illustrated in Figure 8-15, such a relationship was not found in this work. We also studied the variation in resonant frequencies due to surface stresses as compared to the tensile strain in the nanowires, but a similarly inconclusive results was obtained. However, Figure 8-15 does indicate that surface stresses are likely not to strongly impact (more than 2%) the resonant frequencies of fixed/free nanowires unless very small cross sectional areas (< 10 nm) and large aspect ratios are utilized.

8.7.1. Comparison to Experiment

An extensive literature search has revealed that most studies utilizing resonating silicon nanowires involve nanowires with cross sectional sizes generally exceeding

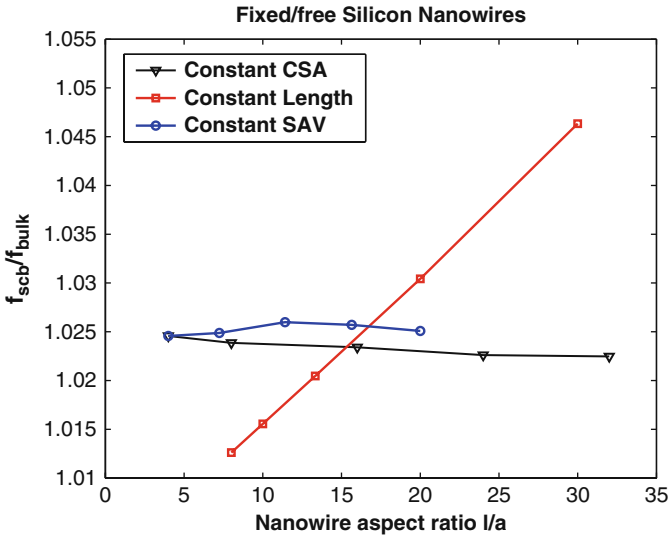


Figure 8-15. Normalized resonant frequencies for fixed/free constant CSA, SAV and length nanowires plotted against the nanowire aspect ratio

50 nm [6,99,30]. At those sizes, both the experimental results and extrapolation of the current SCB results indicate that surface effects will not have a dominant role on the resonant frequencies, and that continuum beam theory should be valid for interpreting the resonant properties.

We did find one study involving the resonant properties of sub-30 nm cross section silicon nanowires, that of Li et al. [29]. The silicon nanowires in that work were of the ⟨110⟩ orientation, and were fabricated in the fixed/fixed configuration. The nanowires were found to have a sharp decrease in Young's modulus, with a 53 GPa Young's modulus reported for 12 nm diameter nanowires. In comparison, using the results in Figure 8-13, we find that the SCB model predicts a 58.5 GPa Young's modulus for a 16 nm cross section ⟨100⟩ nanowire. We note that a direct comparison cannot be made due to the fact that the nanowires in the present work were axially aligned in the ⟨100⟩ direction.

Two other studies involving the mechanical properties of sub-30 nm cross section silicon nanowires were found, with both involving tensile deformation. Kizuka et al. [32] used an AFM to perform tensile elongation of single crystal ⟨100⟩ silicon nanowires with cross sections less than 10 nm; the measured Young's modulus was on the order of 18 GPa, which is considerably smaller than the 90 GPa Young's modulus for bulk ⟨100⟩ silicon.

More recently, Han et al. [31] also performed in situ TEM observation of the tensile failure of ⟨110⟩ silicon nanowires. Nanowire sizes down to 15 nm cross sections were considered; using the activation energy for dislocation nucleation, they also obtained a strong size-dependence in the Young's modulus, with a modulus value

of 55 GPa reported for 15 nm cross section nanowires. Again, the 58.5 GPa modulus obtained using the SCB model for 16 nm cross section ⟨100⟩ nanowires agrees well, though as before, a direct comparison cannot be made due to the different crystallographic orientations.

Despite the small amount of experimental data to which to compare the present results, the present results are qualitatively consistent with available experimental data [29,32,31] and theoretical results [97,100] in predicting a relative decrease in resonant frequencies, and thus Young's modulus, for fixed/fixed nanowires.

8.8. CONCLUSIONS AND PERSPECTIVES

In conclusion, we have presented an overview of the recently developed surface Cauchy-Born model, which is a nonlinearly elastic, finite deformation multiscale continuum model that captures surface stress effects on the mechanical behavior and properties of nanomaterials. The surface Cauchy-Born model leads to a variational framework that enables the solution of nanomechanical boundary value problems for displacements, stresses and strains in nanomaterials using standard nonlinear finite element methods. In this vein, the surface Cauchy-Born model offers the following advantages as compared to the various continuum surface elastic formulations that have been presented [33, 41, 51, 52, 54–57, 101]: (1) No additional atomistic calculations are necessary to obtain surface elastic constants. All surface stresses and elastic constants are obtained directly from the underlying interatomic potential. (2) The surface physics are captured using accurate, multibody interatomic potentials such as the EAM for FCC metals or Tersoff for silicon. (3) The variational form requires no non-standard terms as is common for surface elastic formulations [51], which makes it simple to discretize numerically using standard nonlinear finite element methods. (4) Capturing the atomistic surface stress effects within a continuum framework leads to an enormous reduction in computational expense as compared to fully atomistic calculations while still maintaining accuracy comparable to that achievable in a fully atomistic calculation. (5) No assumption of linear elastic kinematics was required for the surface Cauchy-Born model; instead, the surface Cauchy-Born model is based upon fully nonlinear, finite deformation kinematics.

We first verified the surface Cauchy-Born model in calculating the minimum energy configurations of metal nanowires due to surface stresses. We then utilized the surface Cauchy-Born model to quantify, for the first time, how boundary conditions and surface stresses couple to alter the resonant frequencies of silicon nanowires as compared to those expected using standard continuum beam theory [84]. With regards to the effects of surface stresses on the silicon nanowire resonant frequencies, we have found that: (1) Surface stresses cause significant deviations in the resonant frequencies of nanowires as compared to those that are found using standard continuum beam theory with bulk material properties, with the deviation having a different trend depending on whether fixed/fixed or fixed/free boundary conditions are used. We find that the resonant frequencies of nanowires with cross

sectional lengths greater than about 30 nm show little deviation from those predicted from continuum beam theory. We also find that surface stresses most strongly impact the resonant properties of fixed/fixed silicon nanowires, which are found to decrease substantially as compared to predictions from continuum beam theory. In contrast, surface stresses do not cause substantial deviations from beam theory for fixed/free silicon nanowires unless nanowires with very small cross sectional lengths ($<$10 nm) and large aspect ratios are considered. (2) For fixed/fixed silicon nanowires, accounting for the compressive state of stress resulting from the coupled effects of surface stresses and boundary conditions is critical to capturing the observed reductions in the resonant frequencies as compared to continuum beam theory. (3) The deviation that surface stresses cause in the resonant properties of fixed/fixed nanowires as compared to beam theory scales proportional to the nanowire aspect ratio l/a. (4) No such scaling relationship was found for surface stress effects on the resonant properties of fixed/free nanowires. (5) The present finding that the resonant properties of fixed/fixed silicon nanowires, and therefore the elastic properties such as the Young's modulus decrease with respect to the bulk value qualitatively agrees with recent experimental [29,31,32] and theoretical [97,100] results.

The ability to capture surface effects on the behavior and properties of nanomaterials using accurate and efficient finite element-based models has and will open the door to study other important nanoscale phenomena. To that end, work is already underway to study nanoscale resonant mass sensing [102], surface-driven thermoelastic dissipation in nanomaterials [103], resonant based strain sensing using deformed FCC metal nanowires [104], and the effects of the residual surface stress on the resonant frequencies of metal nanowires if finite deformation kinematics are considered [105]. In general, multiscale models incorporating surface effects on nanomaterials will be invaluable in the near future for their ability to investigate problems that cannot be solved due to computational expense by atomistic calculations alone.

ACKNOWLEDGMENTS

HSP acknowledges NSF grant number CMMI-0750395 in support of this research.

REFERENCES

1. C. M. Lieber, Nanoscale science and technology: building a big future from small things, MRS Bulletin 28(7) (2003) 486–491.
2. P. Yang, The chemistry and physics of semiconductor nanowires, MRS Bulletin 30(2) (2005) 85–91.
3. Y. Xia, P. Yang, Y. Sun, Y. Wu, B. Mayers, B. Gates, Y. Yin, F. Kim, H. Yan, One-dimensional nanostructures: synthesis, characterization, and applications, Advanced Materials 15(5) (2003) 353–389.
4. H. G. Craighead, Nanoelectromechanical systems, Science 290 (2000) 1532–1535.
5. K. L. Ekinci, M.L. Roukes, Nanoelectromechanical systems, Review of Scientific Instruments 76 (2005) 061101.

6. A. N. Cleland, M.L. Roukes, Fabrication of high frequency nanometer scale mechanical resonators from bulk Si crystals, Applied Physics Letters 69(18) (1996) 2653–2655.
7. X. M. H. Huang, C.A. Zorman, M. Mehregany, M.L. Roukes, Nanodevice motion at microwave frequencies, Nature 42 (2003) 496.
8. N. V. Lavrik, M.J. Sepaniak, P.G. Datskos, Cantilever transducers as a platform for chemical and biological sensors, Review of Scientific Instruments 75(7) (2004) 2229–2253.
9. K. L. Ekinci, Electromechanical transducers at the nanoscale: actuation and sensing of motion in nanoelectromechanical systems (NEMS), Small 1(8–9) (2005) 786–797.
10. H. Petrova, J. Perez-Juste, Z.Y. Zhang, J. Zhang, T. Kosel, G.V. Hartland, Crystal structure dependence of the elastic constants of gold nanorods, Journal of Materials Chemistry 16(40) (2006) 3957–3963.
11. E. W. Wong, P. E. Sheehan, C. M. Lieber, Nanobeam mechanics: elasticity, strength, and toughness of nanorods and nanotubes, Science 277 (1997) 1971–1975.
12. B. Wu, A. Heidelberg, J. J. Boland, Mechanical properties of ultrahigh-strength gold nanowires, Nature Materials 4 (2005) 525–529.
13. A. Heidelberg, L. T. Ngo, B. Wu, M. A. Phillips, S. Sharma, T. I. Kamins, J. E. Sader, J. J. Boland, A generalized description of the elastic properties of nanowires, Nano Letters 6(6) (2006) 1101–1106.
14. S. Cuenot, C. Frétigny, S. Demoustier-Champagne, B. Nysten, Surface tension effect on the mechanical properties of nanomaterials measured by atomic force microscopy, Physical Review B 69 (2004) 165410.
15. G. Y. Jing, H. L. Duan, X. M. Sun, Z. S. Zhang, J. Xu, Y. D. Li, J. X. Wang, D. P. Yu, Surface effects on elastic properties of silver nanowires: contact atomic-force microscopy, Physical Review B 73 (2006) 235409.
16. S. Hoffmann, I. Utke, B. Moser, J. Michler, S. H. Christiansen, V. Schmidt, S. Senz, P. Werner, U. Gosele, C. Ballif, Measurement of the bending strength of vapor-liquid-solid grown silicon nanowires, Nano Letters 6(4) (2006) 622–625.
17. Y. Chen, B. L. Dorgan, D. N. McIlroy, D. E. Aston, On the importance of boundary conditions on nanomechanical bending behavior and elastic modulus determination of silver nanowires, Journal of Applied Physics 100 (2006) 104301.
18. T. Namazu, Y. Isono, T. Tanaka, Evaluation of size effect on mechanical properties of single crystal silicon by nanoscale bending test using AFM, Journal of Microelectromechanical Systems 9(4) (2000) 450–459.
19. S. Sundararajan, B. Bhushan, T. Namazu, Y. Isono, Mechanical property measurements of nanoscale structures using an atomic force microscope, Ultramicroscopy 91 (2002) 111–118.
20. A. S. Paulo, J. Bokor, R. T. Howe, R. He, P. Yang, D. Gao, C. Carraro, R. Maboudian, Mechanical elasticity of single and double clamped silicon nanobeams fabricated by the vapor-liquid-solid method, Applied Physics Letters 87 (2005) 053111.
21. M. Tabib-Azar, M. Nassirou, R. Wang, S. Sharma, T. I. Kamins, M. S. Islam, R.S. Williams, Mechanical properties of self-welded silicon nanobridges, Applied Physics Letters 87 (2005) 113102.
22. S. S. Verbridge, J. M. Parpia, R. B. Reichenbach, L. M. Bellan, H. G. Craighead, High quality factor resonance at room temperature with nanostrings under high tensile stress, Journal of Applied Physics 99 (2006) 124304.
23. A. Husain, J. Hone, H. W. C. Postma, X. M. H. Huang, T. Drake, M. Barbic, A. Scherer, M. L. Roukes, Nanowire-based very-high-frequency electromechanical oscillator, Applied Physics Letters 83(6) (2003) 1240–1242.
24. C.Y. Nam, P. Jaroenapibal, D. Tham, D. E. Luzzi, S. Evoy, J. E. Fischer, Diameter-dependent electromechanical properties of GaN nanowires, Nano Letters 6(2) (2006) 153–158.

25. D. A. Dikin, X. Chen, W. Ding, G. Wagner, R. S. Ruoff, Resonance vibration of amorphous nanowires driven by mechanical or electrical field excitation, Journal of Applied Physics 93(1) (2003) 226–230.
26. J. Yang, T. Ono, M. Esashi, Investigating surface stress: surface loss in ultrathin single-crystal silicon cantilevers, Journal of Vacuum Science and Technology B 19(2) (2001) 551–556.
27. B. H. Houston, D. M. Photiadis, M. H. Marcus, J. A. Bucaro, X. Liu, J. F. Vignola, Thermoelastic loss in microscale oscillators, Applied Physics Letters 80(7) (1976) 1300–1302.
28. S. Evoy, A. Olkhovets, L. Sekaric, J. M. Parpia, H. G. Craighead, D. W. Carr, Temperature-dependent internal friction in silicon nanoelectromechanical systems, Applied Physics Letters 77(15) (2000) 2397–2399.
29. X. Li, T. Ono, Y. Wang, M. Esashi, Ultrathin single-crystalline-silicon cantilever resonators: fabrication technology and significant specimen size effect on Young′s modulus, Applied Physics Letters 83(15) (2003) 3081–3083.
30. X. L. Feng, R. He, P. Yang, M. L. Roukes, Very high frequency silicon nanowire electromechanical resonators, Nano Letters 7(7) (2007) 1953–1959.
31. X. Han, K. Zheng, Y .F. Zhang, X. Zhang, Z. Zhang, Z. L. Wang, Low-temperature in situ large-strain plasticity of silicon nanowires, Advanced Materials 19 (2007) 2112–2118.
32. T. Kizuka, Y. Takatani, K. Asaka, R. Yoshizaki, Measurements of the atomistic mechanics of single crystalline silicon wires of nanometer width, Physical Review B 72 (2005) 035333.
33. R. C. Cammarata, Surface and interface stress effects in thin films, Progress in Surface Science 46(1) (1994) 1–38.
34. J. Diao, K. Gall, M. L. Dunn, Surface-stress-induced phase transformation in metal nanowires, Nature Materials 2(10) (2003) 656–660.
35. H. S. Park, K. Gall, J. A. Zimmerman, Shape memory and pseudoelasticity in metal nanowires, Physical Review Letters 95 (2005) 255504.
36. W. Liang, M. Zhou, F. Ke, Shape memory effect in Cu nanowires, Nano Letters 5(10) (2005) 2039–2043.
37. H. S. Park, Stress-induced martensitic phase transformation in intermetallic nickel aluminum nanowires, Nano Letters 6(5) (2006) 958–962.
38. Y. Kondo, K. Takayanagi, Gold nanobridge stabilized by surface structure, Physical Review Letters 79(18) (1997) 3455–3458.
39. Y. Kondo, Q. Ru, K. Takayanagi, Thickness induced structural phase transition of gold nanofilm, Physical Review Letters 82(4) (1999) 751–754.
40. L. G. Zhou, H. Huang, Are surfaces elastically softer or stiffer? Applied Physics Letters 84(11) (2004) 1940–1942.
41. V. B. Shenoy, Atomistic calculations of elastic properties of metallic FCC crystal surfaces, Physical Review B 71 (2005) 094104.
42. H. Liang, M. Upmanyu, H. Huang, Size-dependent elasticity of nanowires: nonlinear effects, Physical Review B 71 (2005) 241403(R).
43. H. S. Park, K. Gall, J. A. Zimmerman, Deformation of FCC nanowires by twinning and slip, Journal of the Mechanics and Physics of Solids 54(9) (2006) 1862–1881.
44. P. Lu, H. P. Lee, C. Lu, S. J. O′Shea, Surface stress effects on the resonance properties of cantilever sensors, Physical Review B 72 (2005) 085405.
45. P. Lu, F. Shen, S. J. O′Shea, K. H. Lee, T. Y. Ng, Analysis of surface effects on mechanical properties of microcantilevers, Materials Physics and Mechanics 4 (2001) 51–55.
46. M. E. Gurtin, X. Markenscoff, R. N. Thurston, Effects of surface stress on the natural frequency of thin crystals, Applied Physics Letters 29(9) (1976) 529–530.
47. J. E. Sader, Surface stress induced deflections of cantilever plates with applications to the atomic force microscope: rectangular plates, Journal of Applied Physics 89(5) (2001) 2911–2921.

48. G. Y. Huang, W. Gao, S. W. Yu, Model for the adsorption-induced change in resonance frequency of a cantilever, Applied Physics Letters 89 (2006) 043506.
49. D. W. Dareing, T. Thundat, Simulation of adsorption-induced stress of a microcantilever sensor, Journal of Applied Physics 97 (2005) 043526.
50. A. W. McFarland, M. A. Poggi, M. J. Doyle, L. A. Bottomley, J. S. Colton, Influence of surface stress on the resonance behavior of microcantilevers, Applied Physics Letters 87 (2005) 053505.
51. M. E. Gurtin, A. Murdoch, A continuum theory of elastic material surfaces, Archives of Rational Mechanics and Analysis 57 (1975) 291–323.
52. R. E. Miller, V. B. Shenoy, Size-dependent elastic properties of nanosized structural elements, Nanotechnology 11 (2000) 139–147.
53. D. E. Segall, S. Ismail-Beigi, T. A. Arias, Elasticity of nanometer-sized objects, Physical Review B 65 (2002) 214109.
54. L. H. He, C. W. Lim, B. S. Wu, A continuum model for size-dependent deformation of elastic films of nano-scale thickness, International Journal of Solids and Structures 41 (2004) 847–857.
55. P. Sharma, S. Ganti, N. Bhate, Effect of surfaces on the size-dependent elastic state of nano-inhomogeneities, Applied Physics Letters 82(4) (2003) 535–537.
56. C. T. Sun, H. Zhang, Size-dependent elastic moduli of platelike nanomaterials, Journal of Applied Physics 92(2) (2003) 1212–1218.
57. R. Dingreville, J. Qu, M. Cherkaoui, Surface free energy and its effect on the elastic behavior of nano-sized particles, wires and films, Journal of the Mechanics and Physics of Solids 53 (2005) 1827–1854.
58. G. Wei, Y. Shouwen, H. Ganyun, Finite element characterization of the size-dependent mechanical behaviour in nanosystems, Nanotechnology 17 (2006) 1118–1122.
59. J. Wang, H. L. Duan, Z. P. Huang, B. L. Karihaloo, A scaling law for properties of nano-structured materials, Proceedings of the Royal Society A 462 (2006) 1355–1363.
60. Z. Tang, H. Zhao, G. Li, N. R. Aluru, Finite-temperature quasicontinuum method for multiscale analysis of silicon nanostructures, Physical Review B 74 (2006) 064110.
61. E. Tadmor, M. Ortiz, R. Phillips, Quasicontinuum analysis of defects in solids, Philosophical Magazine A 73 (1996) 1529–1563.
62. L.E. Shilkrot, R. E. Miller, W. A. Curtin, Multiscale plasticity modeling: coupled atomistics and discrete dislocation mechanics, Journal of the Mechanics and Physics of Solids 52 (2004) 755–787.
63. J. Fish, W. Chen, Discrete-to-continuum bridging based on multigrid principles, Computer Methods in Applied Mechanics and Engineering 193 (2004) 1693–1711.
64. P. A. Klein, J. A. Zimmerman, Coupled atomistic-continuum simulation using arbitrary overlapping domains, Journal of Computational Physics 213 (2006) 86–116.
65. F. F. Abraham, J. Broughton, N. Bernstein, E. Kaxiras, Spanning the continuum to quantum length scales in a dynamic simulation of brittle fracture, Europhysics Letters 44 (1998) 783–787.
66. R. E. Rudd, J. Q. Broughton, Coarse-grained molecular dynamics and the atomic limit of finite elements, Physical Review B 58 (1998) 5893–5896.
67. E. Weinan, Z. Y. Huang, A dynamic atomistic-continuum method for the simulation of crystalline materials, Journal of Computational Physics 182 (2002) 234–261.
68. G. J. Wagner, W. K. Liu, Coupling of atomistic and continuum simulations using a bridging scale decomposition, Journal of Computational Physics 190 (2003) 249–274.
69. H. S. Park, E. G. Karpov, W. K. Liu, P. A. Klein, The bridging scale for two-dimensional atomistic/continuum coupling, Philosophical Magazine 85 (1) (2005) 79–113.
70. H. S. Park, E. G. Karpov, W. K. Liu, A temperature equation for coupled atomistic/continuum simulations, Computer Methods in Applied Mechanics and Engineering 193 (2004) 1713–1732.

71. S.P. Xiao, T. Belytschko, A bridging domain method for coupling continua with molecular dynamics, Computer Methods in Applied Mechanics and Engineering 193 (2004) 1645–1669.
72. W. K. Liu, E. G. Karpov, H. S. Park, Nano Mechanics and Materials: Theory, Multiscale Methods and Applications, John Wiley and Sons, New York, 2006.
73. X. Li, E. Weinan, Multiscale modeling of the dynamics of solids at finite temperature, Journal of the Mechanics and Physics of Solids 53 (2005) 1650–1685.
74. W. K. Liu, E. G. Karpov, S. Zhang, H. S. Park, An introduction to computational nano mechanics and materials, Computer Methods in Applied Mechanics and Engineering 193 (2004) 1529–1578.
75. P.A. Klein, A virtual internal bond approach to modeling crack nucleation and growth, Ph.D. Thesis (1999) Stanford University.
76. M. Arroyo, T. Belytschko, An atomistic-based finite deformation membrane for single layer crystalline films, Journal of the Mechanics and Physics of Solids 50 (2002) 1941–1977.
77. P. Zhang, Y. Huang, P. H. Geubelle, P. A. Klein, K. C. Hwang, The elastic modulus of single-wall carbon nanotubes: a continuum analysis incorporating interatomic potentials, International Journal of Solids and Structures 39 (2002) 3893–3906.
78. E. B. Tadmor, G. S. Smith, N. Bernstein, E. Kaxiras, Mixed finite element and atomistic formulation for complex crystals, Physical Review B 59(1) (1999) 235–245.
79. H. . Park, P. A. Klein, A surface cauchy-born model for silicon nanostructures, Computer Methods in Applied Mechanics and Engineering 197 (2008) 3249–3260.
80. M. S. Daw, M. I. Baskes, Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals, Physical Review B 29(12) (1984) 6443–6453.
81. T. Belytschko, W. K. Liu, B. Moran, Nonlinear Finite Elements for Continua and Structures, John Wiley and Sons, New York, 2002.
82. S. M. Foiles, M. I. Baskes, M. S. Daw, Embedded-atom-method functions for the FCC metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys, Physical Review B 33(12) (1986) 7893–7991.
83. Y. Zhao, B. I. Yakobson, What is the ground-state structure of the thinnest Si nanowires? Physical Review Letters 91(3) (2003) 035501.
84. H.S. Park, Surface stress effects on the resonant properties of silicon nanowires, Journal of Applied Physics 103 (2008) 123504.
85. J. Tersoff, Modeling solid-state chemistry: interatomic potentials for multicomponent systems, Physical Review B 39(8) (1989) 5566–5568.
86. H. Balamane, T. Halicioglu, W. A. Tiller, Comparative study of silicon empirical interatomic potentials, Physical Review B 46(4) (1992) 2250–2279.
87. H. S. Park, P. A. Klein, G. J. Wagner, A surface cauchy-born model for nanoscale materials, International Journal for Numerical Methods in Engineering 68 (2006) 1072–1095.
88. H. S. Park, P. A. Klein, Surface cauchy-born analysis of surface stress effects on metallic nanowires, Physical Review B 75 (2007) 085408.
89. T. J. R. Hughes, The Finite Element Method: Linear Static and Dynamic Finite Element Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1987.
90. J. Diao, K. Gall, M. L. Dunn, Yield asymmetry in metal nanowires, Nano Letters 4(10) (2004) 1863–1867.
91. K. Gall, J. Diao, M. L. Dunn, The strength of gold nanowires, Nano Letters 4(12) (2004) 2431–2436.
92. J. Diao, K. Gall, M. L. Dunn, Atomistic simulation of the structure and elastic properties of gold nanowires, Journal of the Mechanics and Physics of Solids 52 (2004) 1935–1962.
93. Trilinos, http://software.sandia.gov/trilinos/index.html.
94. Tahoe, http://tahoe.ca.sandia.gov.
95. J. Diao, K. Gall, M. L. Dunn, Surface stress driven reorientation of gold nanowires, Physical Review B 70 (2004) 075413.

96. Warp, http://www.cs.sandia.gov/sjplimp/lammps.html.
97. B. Lee, R. E. Rudd, First-principles calculation of mechanical properties of nanowires and comparison to nanomechanical theory, Physical Review B 75 (2007) 195328.
98. W. Weaver, S. P. Timoshenko, D. H. Young, Vibration Problems in Engineering, John Wiley and Sons, New York, 1990.
99. D.W. Carr, S. Evoy, L. Sekaric, H. G. Craighead, J. M. Parpia, Measurement of mechanical resonance and losses in nanometer scale silicon wires, Applied Physics Letters 75(7) (1999) 920–922.
100. J. Q. Broughton, C. A. Meli, P. Vashishta, R. K. Kalia, Direct atomistic simulation of quartz crystal oscillators: bulk properties and nanoscale devices, Physical Review B 56(2) (1997) 611–618.
101. F. H. Streitz, R. C. Cammarata, K. Sieradzki, Surface-stress effects on elastic properties. I. Thin metal films, Physical Review B 49(15) (1994) 10699–10706.
102. G. Yun, H. S. Park, A finite element formulation for nanoscale resonant mass sensing using the surface cauchy-born model, Computer Methods in Applied Mechanics and Engineering 197 (2008) 3324–3336.
103. G. Yun, H. S. Park, A multiscale, finite deformation formulation for surface stress effects on the coupled thermomechanical behavior of nanomaterials, Computer Methods in Applied Mechanics and Engineering 197 (2008) 3337–3350.
104. H. S. Park, Strain sensing through the resonant properties of deformed metal nanowires, Journal of Applied Physics 104 (2008) 013516.
105. H. S. Park, P. A. Klein, Surface stress effects on the resonant properties of metal nanowires: The importance of finite deformation kinematics and the impact of the residual surface stress, Journal of the Mechanics and Physics of Solids 56 (2008) 3144–3166.

CHAPTER 9

PREDICTING THE ATOMIC CONFIGURATION OF 1- AND 2-DIMENSIONAL NANOSTRUCTURES VIA GLOBAL OPTIMIZATION METHODS

C.V. CIOBANU[1], C.Z.WANG[2], D.P.MEHTA[3], AND K.M. HO[2]

[1]*Division of Engineering, Colorado School of Mines, Golden, CO, USA, e-mail: cciobanu@mines.edu;*
[2]*US DOE Ames Laboratory and Iowa State University, Physics Department, Ames, IA, USA, e-mail: wangcz@ameslab.gov; kmh@ameslab.gov*
[3]*Department of Mathematics and Computer Science, Colorado School of Mines, Golden, CO, USA, e-mail: dmehta@mines.edu*

Abstract: In the cluster structure community, global optimization methods are common tools for arriving at the atomic structure of molecular and atomic clusters. The large number of local minima of the potential energy surface of these clusters, and the fact that these local minima proliferate exponentially with the number of atoms in the cluster simply demands the use of fast stochastic methods to find the optimum atomic configuration. Therefore, much of the development work has come from (and mostly stayed within) the cluster structure community. Partly due to wide availability and landmark successes of high resolution microscopy techniques, finding the structure of periodically reconstructed semiconductor surfaces was not posed as a problem of stochastic optimization until recently, when we have shown that high-index semiconductor surfaces can posses a rather large number of local minima with such low surface energies that the identification of the global minimum becomes problematic. We have therefore set out to develop global optimization methods for systems other than clusters, focusing on periodic systems in one- and two- dimensions as such systems currently occupy a central place in the field of nanoscience. In this article, we review some of our recent work on global optimization methods (the parallel-tempering Monte Carlo method and the genetic algorithm) and show examples/results from two main problem categories: (1) the two-dimensional problem of determining the atomic configuration of clean semiconductor surfaces, and (2) finding the structure of freestanding nanowires. While focused mainly on *optimization* the atomic structure for a system with set periodic boundary conditions, our account also reviews a recent example of using genetic algorithms for *growth* of nanostructures into their global energy minima compatible with given confinement conditions

Keywords: Global optimization, Surface, Nanowire

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 231–253.
DOI 10.1007/978-1-4020-9785-0_9,

9.1. INTRODUCTION

The discovery of carbon nanotubes (CNTs) [1] sparked arduous interest in nanotube science and applications. The fascination with nanotubes, now still at an all-time high, is beginning to slow down as researchers have found another class of materials (nanowires) with superior potential to impact science at the nanometer scale, as well as our everyday lives. While CNTs are inert, nanowires (NWs) have "chemistry" which opens up unprecedented avenues for controlling their structure as well as their electronic, optical, mechanical and magnetic properties [2]. The scientific curiosity about nanostructures in general and about NWs in particular comes from the realization that at such small scales the structure, properties, and phenomena cannot be straightforwardly inferred from our knowledge of the bulk forms. The appeal of the NWs is also driven by the continuous miniaturization of electronics and optoelectronics industry, which has achieved the limit in which the interconnection of devices in a reliable and controllable way is particularly challenging. Fervent strides are underway in the preparation of NWs for molecular and nano-electronics applications [2, 3]: such wires (possibly doped or functionalized) can operate both as nanoscale devices and as interconnects [4]. Silicon nanowires (SiNWs) offer, in addition to their appeal as building blocks for nanoscale electronics, the benefit of simple fabrication techniques compatible with the currently well-developed silicon technology.

While remarkable progress has been achieved in the synthesis and device applications of SiNWs, atomic-level knowledge of the structure of ultrathin nanowires remains largely speculative. Stating the obvious, when the diameters are as small as 1 nm, the atomic structure of the NW is the single most important factor that determines its electronic, optical, and mechanical properties, as well as the ensuing phenomena and technological applications. The importance of atomic structure has been emphasized in a sequence of recent high-profile publications [5–10], which bring strongly plausible arguments and simulation evidence for various configurations in certain diameter regimes. The current proposals for SiNW configurations fall into several main categories: fused-fullerenes [5], fused-clathrates [9, 11], diamond structure single-crystals with reconstructed nanofacets [8, 12], polycrystals [7] and high-density phases [10], with each of the categories representing a novelty with respect to previous work. However, when considered together, the works [5–12] appear to collectively suggest that the procedures currently used to investigate the structure of thin nanowires are not reliable. One may be determined to draw this conclusion because seemingly simple questions regarding the SiNW diameter ("what is thinnest stable Si nanowire? " [7]) and its core structure are still under investigation.

The reason for the current situation of the SiNW problem is that this problem is exponentially complex (NP-complete), as we could recognize from a quick comparison with the problem of structure of atomic clusters: for a given number and type of atoms, the only mathematical difference between the two problems (wire vs. cluster) is the periodic boundary condition necessary to simulate the quasi 1-dimensional wires. Yet, in most theoretical approaches to date the structures proposed are not

derived from the kind of robust search procedures that have been employed for atomic clusters. The problem is exacerbated by the fact that in the few-nanometer diameter regime experimental characterization of NWs with atomic resolution is extremely difficult, and thus theoretical proposals for NW structures cannot be easily confirmed or refuted. Since the problem of structure determination is NP complete and the current approaches are not dealing with this aspect, it is likely that a set of structures based on physical intuition (albeit refined using electronic structure relaxations) may not include many low-energy minima which could end up being the global minimum. Even if we assume (purely for the sake of the argument) that *all* the thermodynamically relevant structures of *silicon* NWs have already been reported in previous theoretical studies over the last 7 years, such a solution for the case of silicon will not readily transfer into methodologies or knowledge about NWs made of *other materials*, and we would again have to resort to trying numerous intuitive structures over many years.

Another problem, analog in principle with the above problem of finding the structure of nanowires, is searching for the structure of semiconductor surfaces. Under conditions of ultra-high vacuum, these surfaces reorganize their atomic configuration to minimize the surface energy, and in the process create periodic reconstructions which can repeat almost flawlessly over thousands of Angstroms in the nominal plane of the surface. The determination of atomic structure of crystalline surfaces is a long-standing problem in surface science. Despite major progress brought by experimental techniques such as scanning tunneling microscopy (STM) and advanced theoretical methods for treating the electronic and ionic motion, the commonly used procedures for finding the atomic structure of surfaces still rely to a large extent on one's intuition in interpreting STM images. While these procedures have proven successful for determining the atomic configuration of many low-index surfaces [e.g., Si(001) and Si(111)], in the case of high-index surfaces their usefulness is limited because the number of good structural models for high-index surfaces is rather large, and may not be exhausted heuristically. An illustrative example is Si(5 5 12), whose structure has been the subject of intense dispute [13–16] since the publication of the first atomic model proposed for this surface [17]. There are also other stable surfaces of silicon such as (113) [18, 19] and (105)[20–25], which required a long time for their correct structures to be revealed.

The high-index surfaces attract a great deal of scientific and technological interest since they can serve as natural and inexpensive templates for the fabrication of low-dimensional nanoscale structures. Knowledge about the template surface can lead to new ways of engineering the morphological and physical properties of these nanostructures. The main technique for investigating atomic-scale features of surfaces is STM, although, as pointed out in a recent review, STM alone is only able to provide "a range of speculative structural models which are increasingly regarded as solved surface structures" [26]. The role of theoretical methods for structural optimization of high-index surfaces has been largely reduced to the relaxation of these speculative models. However, the publication of numerous studies that report different structures for a given stable high-index silicon surface (see, e.g., [13–17]) indicates

a need to develop methodologies capable of actually searching for the atomic structure in a way that does not predominantly rely on the heuristic reasoning associated with interpreting STM data. We are thus facing the same problem as described above for the case of nanowires, which is why we develop global optimization methods as useful tools to complement experimental data (when exists or is readily obtainable) or to make more robust predictions of structure in case experiments are not available.

A truly general and robust way of predicting the atomic structure of 1-D and 2-D systems surfaces takes sustained effort over many years. It is not entirely clear that such robust atomic-scale predictions about semiconductor surfaces can even be ventured, since theoretical efforts have been somewhat tempered by the lack of empirical or semiempirical potentials that are *both sufficiently fast and sufficiently transferable* for surface or nanowire calculations.

However, the long process that lead to the discovery of the reconstruction of the (105) surface [20, 27, 22–25] indicates a clear need for a search methodology that does not rely on human intuition. Our efforts to develop such methodologies are presented in Section 2 (parallel-tempering Monte Carlo and the genetic algorithm) for finding the lowest-energy reconstructions for elemental crystal surfaces. We review our work on atomic structures of surfaces and nanowires over the past few years, relying in particular on Refs. [39, 57, 54, 60], from which large portions where reproduced here with permission of the respective publishers. Our initial focus will be on silicon because of its utmost fundamental and technological importance; nonetheless, the same strategies could be applied for any other material surfaces provided suitable models for atomic interactions are available. In Section 9.3 we will show how the genetic algorithm can be modified to address the structure of nanowires, and describe recent results on the magic configurations of the H-passivated SiNWs oriented parallel to the [110] direction. We also review an intriguing application of the genetic algorithms for 1-d nanostructures, namely their growth into global minima compatible with preset confinement conditions. This simulated growth relies on the fact that genetic algorithms for 1-d nanostructures do not have to preserve the number of atoms during the evolution, and can in fact *grow the structure and optimize its atomic configuration during the same run*. Section 9.4 summarizes our results on various surfaces and nanowires obtained so far, and identifies a number of future directions that can tremendously benefit from the application of global search methods.

9.2. RECONSTRUCTION OF SILICON SURFACES AS A PROBLEM OF GLOBAL OPTIMIZATION

In choosing a methodology that can help predict the surface reconstructions, we have taken into account the following considerations. First, the number of atoms in the simulation slab is large because it includes several subsurface layers in addition to the surface ones. Moreover, the number of local minima of the potential energy surface is also large, as it scales roughly exponentially[28, 29] with the number of atoms involved in the reconstruction; by itself, such scaling requires the use of

fast stochastic search methods. Secondly, methods that are based on the modification of the potential energy surface (PES) (such as the basin-hoping[30] algorithm), although very powerful in predicting global minima, have been avoided as our future studies are aimed at predicting not only the correct lowest-energy reconstructions, but also the thermodynamics of the surface. Lastly, the calculation of interatomic forces is expensive, so the method should be based on Monte Carlo algorithms rather than molecular dynamics. We mention, however, that recent advances in molecular dynamics algorithms, especially the parallel replica[31] and temperature accelerated dynamics[32] developed by Voter and coworkers, may constitute viable alternatives to Monte Carlo parallel tempering for the sampling of low-temperature systems.

These considerations, coupled with a desire for simplicity and robustness of implementation, prompted us to choose the parallel-tempering Monte Carlo (PTMC) algorithm [33, 34] for finding the reconstruction of semiconductor surfaces. While we describe the salient features of the PTMC for crystal surfaces in Section 9.2.1, the reader is refer to the original work [35] for full implementation details.

The PTMC simulations, however, have a broader scope that the global minimum search, as they are used to perform a thorough thermodynamic sampling of the surface systems under study. Given their scope, such calculations [35] are very demanding, usually requiring several tens of processors that run canonical simulations at different temperatures and exchange configurations in order to drive the low-temperature replicas into the ground state. If we focus only on finding the reconstructions at zero Kelvin (which can be representative for crystal surfaces in the low-temperature regimes achieved in laboratory conditions), it is then justified to explore alternative methods for finding the structure of high-index surfaces. In Section 9.2.2, we will address the problem of surface structure determination at zero Kelvin, and report a genetically-based strategy for finding the reconstructions of elemental semiconductor surfaces. Our choice for developing this genetic algorithm (GA) was motivated by its successful application for the structural optimization of atomic clusters [36, 37]. We have designed and tested the algorithm for Si(105)[38], but we tested it on other surfaces as well [39–41]. Both Sections 9.2.1. and 9.2.2. deal with the Si(114)-(2 $\times$ 1) surface, as an illustrative example of how the two methodologies fare in the quest for finding low energy structures.

9.2.1. The Parallel-Tempering Monte Carlo

The reconstructions of semiconductor surfaces are determined not only by the efficient bonding of the surface atoms, but also by the stress created in the process [17]. Therefore, we retain a large number of subsurface atoms when performing a global search for the lowest energy configuration: this way the surface stress is intrinsically considered when reconstructions are sorted out. The number of local minima of the potential energy is also large, as it scales roughly exponentially [28, 29] with the number of atoms involved in the reconstruction; by itself, such scaling requires the use of fast stochastic search methods. One such method is the parallel-tempering Monte Carlo (PTMC) algorithm [33, 34], which was shown to successfully find the

reconstructions of a vicinal Si surface when coupled with an exponential cooling [35]. Before outlining the procedure, we discuss briefly the computational cell and the empirical potential used.

The simulation cell [of dimensions $3a \times a\sqrt{2}$ for Si(114)] in the plane of the surface) has a single-face slab geometry with periodic boundary conditions applied in the plane of the surface, and no periodicity in the direction normal to it. The "hot" atoms from the top part of the slab (10–15 Å thick) are allowed to move, while the bottom layers of atoms are kept fixed to simulate the underlying bulk crystal. The area of the simulation cell and the number of atoms in the cell are kept fixed during each simulation. Under these conditions, the problem of finding the most stable reconstruction reduces to the global minimization of the total potential energy $V(\mathbf{x})$ of the atoms in the simulation cell (here $\mathbf{x}$ denotes the set of atomic positions). In terms of atomic interactions, we are constrained to use empirical potentials because the highly accurate ab-initio or tight-binding methods are prohibitive as far as the search itself is concerned. Since this work is aimed at finding the *lowest* energy reconstructions for arbitrary surfaces, the choice of the empirical potential is important. After numerical experimentation with several empirical models, we chose to use the highly optimized empirical potential (HOEP) recently developed by Lenosky et al. [42]. HOEP is fitted to a large database of ab-initio calculations using the force-matching method, and provides a good description of the energetics of all atomic coordinations up to $Z = 12$.

The parallel tempering Monte Carlo method (also known as the replica-exchange Monte-Carlo method) consists in running parallel canonical simulations of many statistically independent replicas of the system, each at a different temperature $T_1 < T_2 < \ldots < T_N$. The set of N temperatures $\{T_i,\ i = 1,2,\ldots,N\}$ is called a temperature schedule (or schedule for short). The probability distributions of the individual replicas are sampled with the Metropolis algorithm [43], although any other ergodic strategy can be employed [44]. Irrespective of what sampling strategy is being used for each replica, the key feature of the parallel tempering method is that swaps between replicas of neighboring temperatures T_i and T_j ($j = i \pm 1$) are proposed and allowed with the conditional probability [33, 34] given by

$$\min\left\{1, e^{(1/T_j - 1/T_i)[V(\mathbf{x}_j) - V(\mathbf{x}_i)]/k_B}\right\}, \tag{9-1}$$

where $V(\mathbf{x}_i)$ represents the energy of the replica i and k_B is the Boltzmann constant. The conditional probability (9-1) ensures that the detailed balance condition is satisfied and that the equilibrium distributions are the Boltzmann ones for each temperature.

In the limit of low temperatures, the PTMC procedure allows for a geometric temperature schedule [45, 46]. To show this, we note that when the temperature drops to zero, the system is well approximated by a multidimensional harmonic oscillator, so the acceptance probability for swaps attempted between two replicas with temperatures $T < T'$ is given by the incomplete beta function law [46]

$$Ac(T,T') \simeq \frac{2}{B(d/2,d/2)} \int_0^{1/(1+R)} \theta^{d/2-1}(1-\theta)^{d/2-1} d\theta, \tag{9-2}$$

where d denotes the number of degrees of freedom of the system, B is the Euler beta function, and $R \equiv T'/T$. Since it depends only on the temperature ratio R, the acceptance probability (9-2) has the same value for any arbitrary replica running at a temperature T_i, provided that its neighboring upper temperature T_{i+1} is given by $T_{i+1} = RT_i$. The value of R is determined such that the acceptance probability given by Eq. (9-2) attains a prescribed value p. Thus, the (optimal) schedule that ensures a constant probability p for swaps between neighboring temperatures is a geometric progression:

$$T_i = R^{i-1} T_\mathrm{min}, \quad 1 \leq i \leq N, \tag{9-3}$$

where $T_\mathrm{min} = T_1$ is the minimum temperature of the schedule.

The typical Monte Carlo simulation done in this work consists of two main parts that are equal in terms of computational effort. In the first stage of the computation, we perform a parallel tempering run for a range of temperatures $[T_\mathrm{min}, T_\mathrm{max}]$. The configurations of minimum energy are retained for each replica, and used as starting configurations for the second part of the simulation, in which replicas are cooled down exponentially until the largest temperature drops below a prescribed value. As a key feature of the procedure, the parallel tempering swaps are not turned off during the cooling steps. Thus, in the second part of the simulation we are in fact using a combination of parallel tempering and simulated annealing, rather than a simple cooling. At the k-th cooling step, each temperature from the initial temperature schedule $\{T_i, i = 1,2,\ldots,N\}$ is decreased by a factor which is independent of the index i of the replica, $T_i^{(k)} = \alpha_k T_i^{(k-1)}$. Because the parallel tempering swaps are not turned off, we require that at any cooling step k all N temperatures must be modified by the same factor α_k in order to preserve the original swap acceptance probabilities. We have used a cooling schedule of the form [35]

$$T_i^{(k)} = \alpha T_i^{(k-1)} = \alpha^{k-1} T_i \quad (k \geq 1), \tag{9-4}$$

where $T_i \equiv T_i^{(1)}$ and $\alpha = 0.85$.

The third and final part of the minimization procedure is a conjugate-gradient optimization of the last configurations attained by each replica. The relaxation is necessary because we aim to classify the reconstructions in a way that does not depend on temperature, so we compute the surface energy at zero Kelvin for the relaxed slabs i, $i = 1,2,\ldots,N$. The surface energy γ is defined as the excess energy (with respect to the ideal bulk configuration) introduced by the presence of the surface:

$$\gamma = (E_\mathrm{m} - n_\mathrm{m} e_\mathrm{b})/A \tag{9-5}$$

where E_{m} is the potential energy of the n_{m} atoms that are allowed to move, $e_{\mathrm{b}} = -4.6124\,\mathrm{eV}$ is the bulk cohesion energy given by HOEP, and A is the surface area of the slab.

At the end of the simulation, we analyze the energies of the relaxed replicas. Typical plots showing the surface energies of the structures retrieved by the PTMC replicas are shown in Figure 9-1a, for different numbers of particles in the computational cell. To exhaust all the possibilities for the numbers of particles corresponding to the supercell dimensions of $3a \times a\sqrt{2}$, we repeat the PTMC simulation for different values of n ranging from 208 to 220, and look for a periodic behavior of the lowest surface energy as a function of n. For the case of Si(114), this periodicity occurs at intervals of $\Delta n = 4$, as shown in Figure 9-1b. Therefore, the (correct) number of atoms n at which the lowest surface energy is attained is $n = 216$, up to an integer multiple of Δn. As we shall show in Section 9.2.2 below, the repetition of

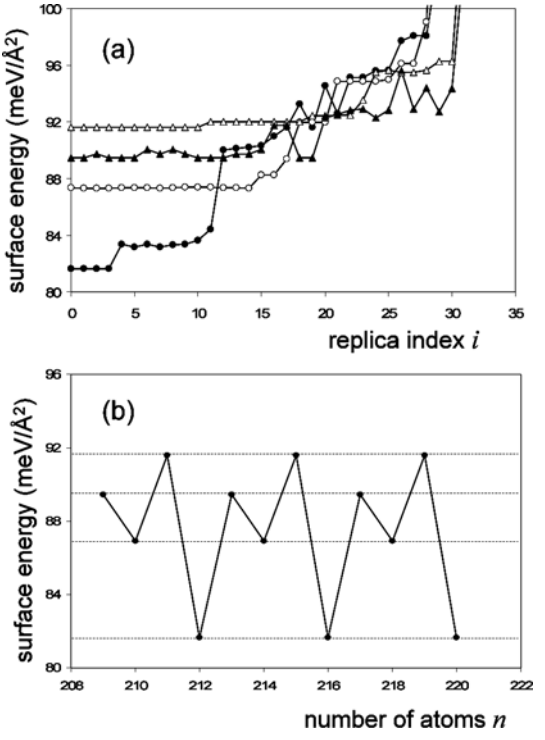


Figure 9-1. (**a**) Surface energies of the relaxed parallel tempering replicas i, ($0 \leq i \leq 31$) with total number of atoms $n = 216$ (*solid circles*), 215 (*open triangles*), 214 (*open circles*) and 213 (*solid triangles*). For clarity, the range of plotted surface energies was limited from above at $100\,\mathrm{meV/\AA^2}$. (**b**) Surface energy of the global minimum structure showing a periodic behavior as a function of n, with a period of $\Delta n = 4$; this finding helps narrowing down the set of values for n that need to be considered for determining the Si(114) reconstructions that have a $3a \times a\sqrt{2}$ periodic cell. (Reproduced from Ref. [39], with permission from Elsevier)

the simulation for different values of n in the simulation cell can be avoided within a genetic algorithm approach.

9.2.2. Genetic Algorithm

Like the previous method, the genetic algorithm also circumvents the intuitive process when proposing candidate models for a given high-index surface. An advantage of this algorithm over most of the previous methodologies used for structural optimization is that the number of atoms involved in the reconstruction, as well as their most favorable bonding topology, can be found within the same genetic search [38].

This search procedure is based on the idea of evolutionary approach in which the members of a generation (pool of models for the surface) mate with the goal of producing the best specimens, i.e. lowest energy reconstructions [38]. "Generation zero" is a pool of p different structures obtained by randomizing the positions of the topmost atoms (thickness d), and by subsequently relaxing the simulation slabs through a conjugate-gradient procedure. The evolution from a generation to the next one takes place by mating, which is achieved by subjecting two randomly picked structures from the pool to a certain operation (mating) $\mathcal{O}$:(A,B) $\longrightarrow$ C. The mating operation $\mathcal{O}$ produces a child structure C from two parent configurations A and B, as follows. The topmost parts of the parent models A and B (thickness d) are separated from the underlying bulk and sectioned by an arbitrary plane perpendicular to the surface. The (upper part of the) child structure C is created by combining the part of A that lies to the left of the cutting plane and the part of slab B lying to the right of that plane: the assembly is placed on a thicker slab, and the resulting structure C is subsequently relaxed.

A mechanism for the survival of the fittest is implemented as a defining feature of the genetic evolution. In each generation, a number of m mating operations are performed. The resulting m children are relaxed and considered for the possible inclusion in the pool based on their surface energy. If there exists at least one candidate in the pool that has a higher surface energy than that of the child considered, then the child structure is included in the pool. Upon inclusion of the child, the structure with the highest surface energy is discarded in order to preserve the total population p. As described, the algorithm favors the crowding of the ecology with identical metastable configurations, which slows down the evolution towards the global minimum. To avoid the duplication of members, we retain a new structure only if its surface energy differs by more than δ when compared to the surface energy of any of the current members p of the pool. Relevant values for the parameters of the algorithm are given in [38]: $10 \leq p \leq 40$, $m = 10$, $d = 5$Å , and $\delta = 10^{-5}$ meV/ Å^2.

We have developed two versions of the algorithm. In the first version, the number of atoms n is kept the same for every member of the pool by automatically rejecting child structures that have different numbers of atoms from their parents (mutants). In the second version of the algorithm, this restriction is not enforced, i.e. mutants are allowed to be part of the pool: in this case, the procedure is able to select the

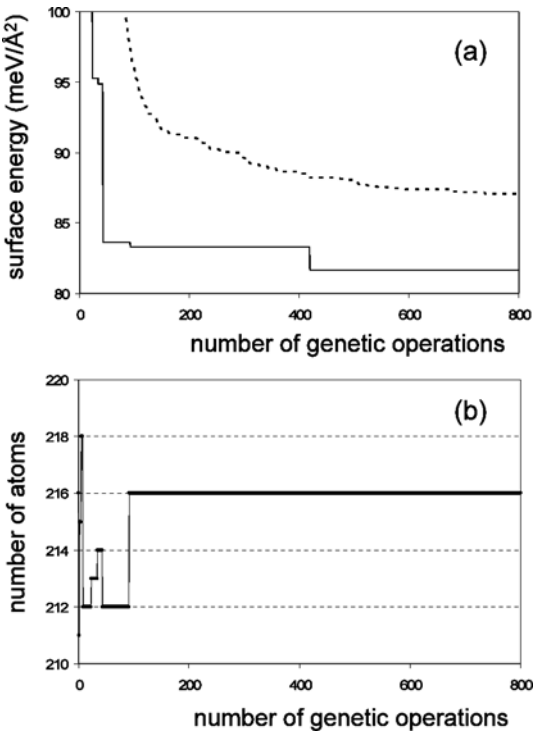


Figure 9-2. (**a**) Evolution of the lowest surface energy (*solid line*) and the average energy (*dash line*) for a pool of $p = 30$ structures during a genetic algorithm (GA) run with variable n ($210 \leq n \leq 222$). (**b**) Evolution of the number of atoms n that corresponds to the model with the lowest energy from the pool, during the same GA run. Note that the lowest energy structure of the pool spends most of its evolution in states with numbers of atoms that are compatible with the global minimum, i.e. $n = 212$ and $n = 216$ (Reproduced from [39], with permission from Elsevier)

correct number of atoms for the ground state reconstruction without any increase over the computational effort required for one single constant-n run. The results of a variable-n run are shown in Figure 9-2a which shows how the lowest energy and the average energy from a pool of $p = 30$ structures decreases as the genetic algorithm run proceeds. The plot in Figure 9-2a displays typical features of the evolutionary approach: the most unfavorable structures are eliminated from the pool rather fast (initial steep transient region of the graphs) and a longer time is taken for the algorithm to retrieve the most stable configuration. The lowest energy structure is retrieved in less than 500 mating operations. The correct number of atoms [refer to Figure 9-2b] is retrieved much faster, within approximately 100 operations. It is worth noting that even during the transient period, the lowest-energy member of the pool spends most of its evolution in a state with a number of atoms ($n = 212$) that is compatible with the global minimum structure.

The two independent algorithms (PTMC and GA) presented here are able to retrieve a set of possible candidates for the lowest energy surface structure. We use

both of the algorithms in this work in order to assess how robust their structure predictions are. As it turns out, the two methods not only find the same lowest energy structures for each value of the total number of atoms n, but also most of the other low-energy reconstructions – a finding that builds confidence in the quality of the configuration sampling performed here. Since the atomic interactions are modelled by an empirical potential [42], it is desirable to check the relative stability of different model structures using higher-level calculations based on density functional theory (DFT). The details of these density functional calculations can be found, e.g., in [39].

9.2.3. Selected Results on Si(114)

At the end of the global search procedures (PTMC and GA), we obtain a set of model structures which we sort by the number of atoms in the simulation cell and by their surface energy. Since the empirical potentials may not be fully transferable to different surface environments, we study not only the global minima given by the model for different values of n, but also most of the local minima that are within 15 meV/ Å^2 from the lowest energy configurations. After the global optimizations, the structures obtained are also relaxed by density functional theory (DFT) methods [39, 40]. The results are summarized in Table 9-1 below.

Table 9-1. Surface energies of selected Si(114) reconstructions, sorted by the number of atoms n in the $3a \times a\sqrt{2}$ periodic cell. The second column shows the number of dangling bonds (counted for structures relaxed with HOEP) per unit area. The last two columns list the surface energies given by the HOEP interaction model [42] and by density functional calculations (DFT) [47] with the parameters described in text. (Adapted from [39] with permission from Elsevier)

n	Bond counting ($db/3a^2\sqrt{2}$)	HOEP (meV/ Å^2)	DFT (meV/ Å^2)
216	8	81.66	89.48
	8	83.16	90.34
	8	83.31	91.29
	8	83.39	88.77
	8	83.64	94.68
	8	84.42	92.16
215	8	91.61	97.53
	8	91.82	95.30
	8	92.00	94.20
214	6	86.95	95.17
	10	87.32	99.58
	10	87.39	98.47
	10	87.49	93.88
213	4	89.46	90.43
	6	89.76	94.01

Table 9-1 lists the density of dangling bonds (db per area), as well as the surface energies of several different models calculated using the HOEP potential and DFT. The configurations have been listed in increasing order of the surface energies computed with HOEP, as this is the actual outcome of the global optimum searches. The data shows clearly that the density of dangling bonds at the Si(114) surface is, in fact uncorrelated with the surface energy. The lowest number of dbs per area reported here is 4, and it corresponds to $n = 213$ and $\gamma = 90.43$ meV/ Å^2 at the DFT level. The optimum structure, however, has twice as many dangling bonds but its surface energy is smaller, 88.77 meV/ Å^2. Furthermore, for the same number of atoms in the supercell ($n = 216$) and the same dangling bond density ($8db/3a^2\sqrt{2}$), the different reconstructions obtained via global searches span an energy interval of at least 5 meV/ Å^2. These findings constitute a clear example that the number of dangling bonds can not be used as a criterion for selecting model reconstructions for Si(114). We expect this conclusion to hold for many other high-index semiconductor surfaces as well.

The HOEP surface energy and the DFT surface energy also show very little correlation, indicating that the transferability of the interaction model [42] for Si(114) is not as good as, for instance, in the case of Si(001) and Si(105) [35]. The most that can be asked from this model potential [42] is that the observed reconstruction [50] is amongst the lower lying energetic configurations – which, in this case it is. We have also tested the transferability of HOEP for the case of Si(113), and found that, although the ad-atom interstitial models [18] are not the most stable structures, they are still retrieved by HOEP as local minima of the surface energy. We found that the low-index (but much more complex) Si(111)-(7 × 7) reconstruction is also a local minimum of the HOEP interaction model, albeit with a very high surface energy. Other tests indicated that, while the transferability of HOEP to the Si(114) orientation is marginal in terms of sorting structural models by their surface energy, this potential [42] performs much better than the more popular interaction models [48, 49], which sometimes do not retrieve the correct reconstructions even as local minima. Therefore, HOEP is very useful as a way to find different local minimum configurations for further optimization at the level of electronic structure calculations.

A practical issue that arises when carrying out the global searches for surface reconstructions is the two-dimensional periodicity of the computational slab. In general, if a periodic surface pattern has been observed, then the lengths and directions of the surface unit vectors may be determined accurately through experimental means (e.g., STM or LEED analysis): in those cases, the periodic vectors of the simulation slab should simply be chosen the same as the ones found in experiments. When the surface does not have two-dimensional periodicity, or when experimental data is difficult to analyze, then one should systematically test computational cells with periodic vectors that are integer multiples of the unit vectors of the bulk truncated surface, which are easily computed from knowledge of crystal structure and surface orientation. There is no preset criterion as to when the incremental testing of the size of the surface cell should be stopped – other than the limitation imposed by

finite computational resources; nevertheless, this approach gives a systematic way of ranking the surface energies of slabs of different areas, and eventually finding the global minimum surface structure.

In this section we have reviewed the PTMC and GA methods of global optimization, and exemplified their application using the case case of Si(114), a stable high-index orientation of silicon. The PTMC and GA procedures coupled with the use of a highly optimized interatomic potential for silicon have lead to finding a set of possible models for Si(114), whose energies have been recalculated via ab initio density functional methods. The most stable structure obtained here without experimental input coincides with the structure determined from scanning tunneling microscopy experiments and density functional calculations in [50]. Motivated by these results for 2-dimensional systems, we have set out to adapt the less computationally intensive of the two methods (GA) for the study of quasi-1-dimensional nanowire systems. The blue print of this genetic algorithm for 1-d systems, along with an example application for the case of hydrogenated SiNWs is presented in the next section.

9.3. THE STRUCTURE OF FREESTANDING NANOWIRES

Interestingly, the application of GA for the study of nanowires has been around for the last few years, and as we have found one research group and their close collaborators [51–53] who used GA for finding the structure of metallic nanowires. To the best of our determination from [51–53], the method can be traced to the original article of Deaven and Ho on molecular clusters [36], since very little has been reported in terms of actual GA procedure for nanowires. While the spirit of the algorithm is general in that it is based on the natural evolution of living ecosystems, the adaptation of GA for artificial systems of atoms requires a certain amount of inspiration, design, as well as intense testing of the versatility of the genetic operations (cross-overs, mutations) for specific boundary conditions. Once GA is fully set up for 1-dimensional boundary conditions (possibly including variable unit-cell period and variable numbers of atoms), then it can be used for NWs made of any material provided that suitable (i.e. fast and sufficiently accurate) atomic interaction models are available. In what follows we describe the blueprint of such algorithm (Section 9.3.1) and show its application for hydrogenated SiNWs (Section 9.3.2).

9.3.1. A Genetic Algorithm for 1-D Nanowire Systems

As in the 2-dimensional case, the GA uses concept of a genetic pool to search for low energy structures using principles inspired by the evolution of biological systems. In biological evolution, offspring generations inherit traits from the older generations that may or may not help them survive. Similarly, during GA simulations, new NW configurations (children) inherit diverse structural motifs (genes) from their "parent" structures that may lower their formation energies, case in which the NW children live. It may also happen that new NW structures have too high energies to "survive", case in which they would not be accepted in the genetic pool. We have shown

above an example of how GA can be used to find the structure of 2-dimensional reconstructions. The key to extending the use of GA to NWs is to realize that these algorithms must be system-specific. In other words, the type of genetic moves (e.g. mating operations) that define the evolution cannot be identical for e.g., clusters [36] and surfaces [38]. On the other hand, the moves and their sequence should not be so radically different that each new physical system would require a prohibitive amount of programming and testing.

In Section 9.2.2. we described a simple GA scheme in 2-D where the main assumptions were that the periodic lengths are a priori known, and that the number of Si atoms in the supercell was either fixed or left to vary. If we consider the vast number of proposals for SiNWs [5–10], it becomes apparent that an algorithm with constant periodic length cannot sort through such multitude of structures with different periods along the wire. We therefore describe here a GA with variable periodic length and variable number of atoms (e.g., with fewer than 100 atoms per unit cell), so that highly unfavorable numbers of atoms can be eliminated quickly and naturally during the genetic evolution. It is our hope that such implementation of GA will prove very versatile as it will be able to simultaneously find the number, the period and the atomic structure of thin NW structures.

We now describe the salient features of the proposed GA. We again start with a pool of p structures, in which the atoms are placed at random (but connected) positions. The key modification is that we allow the boundaries of the periodic cells to relax along with the atomic positions. The "Generation Zero" itself (see Section 9.2.2) will have members of different numbers of atoms and different periodic lengths. The evolution proceeds as follows (refer to Figure 9-3). Two members are randomly chosen from the pool (parent structures a and b), scaled to have the same length (e.g., the geometric mean of their periodic lengths) and translated so that they lie between the same spatial bounds in the periodic direction (z) with their centers of mass projecting in the same location in a plane perpendicular to z. a child

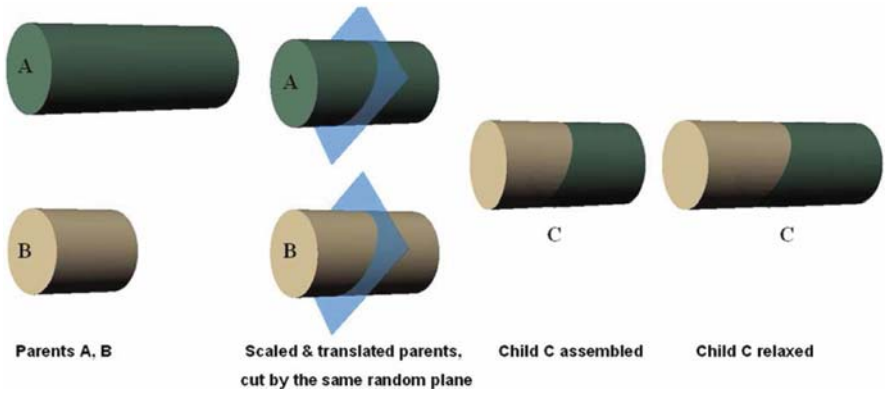


Figure 9-3. Schematics of the cross-over operation in a genetic algorithm design to tackle variations of the periodic cell along its axis. Reproduced from [54], with permission from Taylor & Francis

structure c is produced from a and b through a cross-over (mating) operation in which the parents are sectioned by the same random plane (Figure 9-3) then portions of parents that lie on different sides of the cutting plane are assembled to form C. There are two possibilities to form the child structure, and one of them will be discarded at random without an energy calculation. The child structure c is relaxed (both atomic positions and periodic boundaries), and placed in the pool if its total energy per atom is smaller than that of the most unfavorable pool member. Upon placing the structure in the pool, the highest energy structure is removed in order to preserve the total population p. The procedure is repeated for thousands of generations and stopped according to a set of desired criteria. Since we are interested in diverse types of low energy structures, we would not want the pool crowded with the same structure. For this reason, we allow a child to enter the pool not only when its formation energy is favorable, but also if its structure is different from all the other structures in the pool.

The boundary conditions. The periodicity can still be fixed from the start when it is known from experiments [55] or when we attempt to study wires with crystalline core [56]. In the latter case, one subtlety arises as we need to consider integer multiples of the period of the core in order to allow for various reconstructions to form along the wire [8]. In the case of moving boundary conditions, we will have to ensure from the beginning that our system is connected, as otherwise connectivity may be hard to achieve during the GA evolution.

The cross-over operations. In its most simple form, the cross-over operation is the one shown in Figure 9-3 We can readily imagine other ways to perform cross-over operations such as using two random planes to cut and assemble parts of two or more parents. Based on extensive testing for the high-index surfaces [38–40], we found that cross-over operations with 2 parents and either 1 or 2 cutting planes are sufficiently robust and efficient for a wide range of aspect ratios of the structures, and there is little need to go beyond these values for the thickness regimes that we are interested in. For the crystalline-core nanowires the planes need not be random, but rather parallel to the axis of the nanowire in order for the supercell boundaries not to act like cutting planes themselves. This compatibility between periodic boundary conditions and the mating operations only appears necessary for wires with known core crystal structures: in all other cases it may negatively affect the course of the evolution as the parents would be limited in terms of the kind of offspring they could produce using cutting planes parallel to the axis.

The mutations. As described, there are no explicit mutations in the algorithm. One can envision the simplest standard mutation as selecting one atoms and moving it arbitrarily by a small distance. Many of these standard mutations will have to be forcefully accepted to diversify in the pool, since random mutations almost always increase the energy per atom. More efficient moves that are aimed at improving the diversity of the parents participating in mating operations are zero-penalty, global "mutations" that amount to rotating a child around its axis, shuffling its atomic coordinates through boundary conditions, or taking the mirror image of the child with respect to a plane perpendicular to its axis.

9.3.2. Magic Structures of H-Passivated Si-[110] Nanowires

Our choice to first study this particular NW system (i.e. hydrogenated SiNW oriented along the [110] direction) was motivated by recent experiments [55] that succeeded in characterizing, with atomic-scale resolution, H-SiNWs with diameters between 2 and 7 nm. The authors of [55] have showed that the H-SiNWs are single-crystals with axis orientations along [110], [112] or [111], and they have also reported STM imaging of the NW facets. Comparison with these experiments leaves therefore little wiggle room for theory, and thus constitutes a solid testing bed for our approach. It should be mentioned that in experiments what is being controlled is the number of H atoms on the surface and the wire diameter through timed exposure to the HF environment. What we control in our simulations is the chemical potential of H atoms, and determine the most stable thermodynamic state for a given chemical potential [57].

During a GA optimization run, a pool of at $p = 60$ structures (initially just random collections of atoms with periodic boundary conditions with a fixed period of 3.84 Å, corresponding to the [110] crystal axis) is evolved by performing genetic (mating) operations. For this particular SiNW system, the mating operations consist in selecting two random parent structures from the pool, cutting them with the *same plane parallel to the wire axis*, then combining parts of the parent structures that lie on the opposite sides on the cutting plane to create a new structure (child). The child structure is then passivated by satisfying all its dangling bonds with H atoms, then relaxed with the Hansel-Vogl (HV) empirical model [58]. We include the child structure in the genetic pool based on its formation energy f defined as

$$f = (E - \mu_{\mathrm{H}} n_{\mathrm{H}})/n - \mu, \tag{9-6}$$

where E is the total energy of the computational cell with n Si atoms and n_{H} hydrogen atoms, μ is the bulk cohesive energy of Si in its diamond structure, and μ_{H} is the chemical potential of hydrogen. The H chemical potential is set such that certain hydrogenation reactions at surfaces are thermodynamically allowed [57]. The pool is divided into two equal subsets, one for each values of H. The mating operations are performed both with parents in the same subset and with parents in different subsets, in order to ensure a superior sampling of the potential energy landscape. The mating operation is carried out 15 times during a generation, and a typical GA run has 50,000 generations. At the end of each run, all structures are relaxed at DFT level using the VASP package [59]. The chemical potential H used to compute the DFT formation energies is determined so that it maximizes the correlation with the HV formation energies for a few hundred configurations [57]. While some energetic reordering does occur after the DFT calculations, most of the low-energy structures found with the HV model remain relevant at the DFT level. GA runs with numbers of atoms in the range $9 < n < 31$ revealed three classes of spatially closed structures with relatively low formation energies, which we described as 6-atom-ring chains, double-chains (fused pairs of 6-ring chains) and hexagons (refer to Figure 9-4). As the number of atoms per unit of wire length increases, we found that the most stable

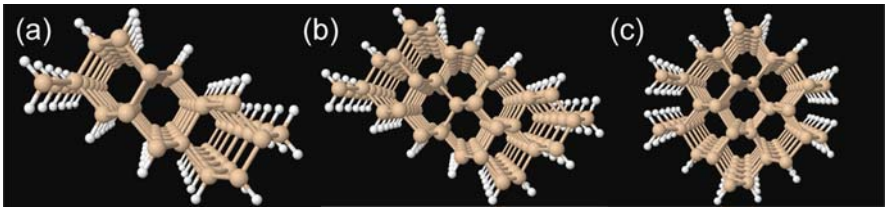


Figure 9-4. Magic nanowires (perspective view) found as minima of the formation energy per atom, Eq. (9-6): the crossection of the wire can be (**a**) chain, (**b**) double-chain, or (**c**) hexagonal

cross section of the nanowire evolves from chains of six-atom rings to double-chains to hexagons bounded by 001 and 111 facets. Our calculations predict that hexagonal wires become stable starting at about 1.2 nm diameter, which is consistent with recent experiments on NWs with diameters of about 3 nm [55]. Pursuing further the comparison with [55], we computed STM images for the facets of hexagons with diameters in the range of 2–3 nm. Our calculation is in agreement with the STM experiments, which also showed the exclusive presence of dihydride species on the 001 facets of [110] H-SiNWs.

9.3.3. Growth of 1-D Nanostructures into Global Minima Under Radial Confinement

In this subsection we present a novel application of genetic algorithms, namely their use for finding the structure of a 1-D nanotube *via simulated growth* [60]. Because we pursue simulation of growth [60], we have chosen a *variable-number* genetic algorithm which in the case of 2-D and 3-D periodic structures was shown to retrieve the correct global minimum of the relevant energetic quantity – i.e., surface energy in 2-D,[38, 39] and cohesion energy per particle in 3-D[61]. The 1-D nanotube studies reviewed here [60] were started with very few atoms in the periodic cell, often with a single atom. The growth of the nanostructures takes off and proceeds entirely through crossover operations, and stops when the optimal structure (i.e., that with lowest energy per particle) for the given confinement conditions is found. As such, the growth is not a reflection of the kinetic processes that occur in actual synthesis experiments, but it is rather a different way to seek the *optimal* nanostructure that can be synthesized under the prescribed confinement conditions.

We have noticed that simple planar crossovers such as those shown in Figure 9-3 lead to extremely slow and often nonconvergent genetic evolutions. Therefore, we have diversified the set of mating (crossover) operations in order to improve the algorithm's performance. Two types of crossover operations were employed. The first type, which we call *sine* crossovers (S_1, S_2, S_3 cf. [60]) requires cutting the parents along sinusoidal lines that are compatible with the periodic boundary conditions along the axis of the tube and with respect to the angular coordinate. This crossover procedure is adapted from recent work in 3-D crystal structure prediction [61] which showed that, at least for 3-D periodic systems, the real-space GA is more

efficient when using cutting functions that obey the periodic boundary conditions. The other type of crossover is based on planar cuts (P_1 and P_2 cf. Ref. [60]) where planes are either randomly oriented or are parallel to the axis of the nanostructure.

We have tested the genetic algorithm for two systems, LJ systems and carbon nanotubed subjected to radial confinement conditions [60], and describe here only the latter system. The purpose is to find out if nanotubes can grow via genetic operations. To this end, we started the GA runs from genetic pools in which each member has one single carbon atom in the periodic cell. We follow the evolution of GA runs in which only one type of cross-over is employed from the set , as well as GA runs where *all* the crossover operations are attempted *with equal probability*. The evolution of the lowest energy in the pool for each of the six GA runs (i.e. five runs based on a single type of crossover, and one run with all equiprobable crossovers) is plotted in Figure 9-5a, and the average energy across the pool is

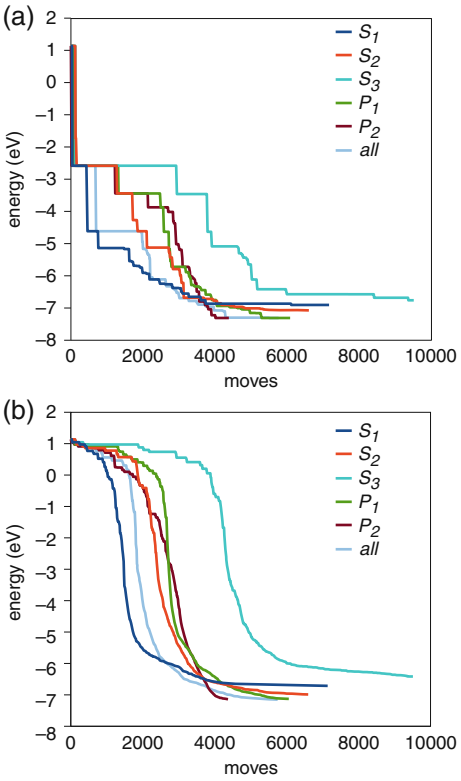


Figure 9-5. Evolution of (**a**) the lowest energy in the pool and (**b**) the average energy across the pool for CNT systems in separate runs performed with only one type of crossover, as well as in a run performed with all crossovers attempted with equal probability. The *horizontal axis* of each plot shows the number of crossovers (moves) attempted. Note that the GA runs based solely on sine operations could *not* find the global minimum structure within 10,000 crossovers. Reproduced from [60], with permission from Taylor & Francis

shown in Figure 9-5b for each run. We observe that only the runs based on planar cuts (P_1 and P_2) are able to find the optimal defect free CNT structure in less than 10^4 operations. The run based on all operations finds the correct structure in less than 5000 moves (crossovers). Since the individual sine crossovers are slow in finding the global optimum structure, the success of the all-move GA run is most likely due to the planar crossovers.

The structure of the best member in the genetic pool for the all-crossovers GA run on CNT systems is shown at selected time-points in the evolution in Figure 9-6. As specified, the run started with a single carbon atom for every member of the pool, but we show two unit cells for each frame to help the visualization through periodic boundary conditions along the direction of the nanotube. The best structure grows from one atom to a string of atoms that spans the length of the periodic cell (frame labeled 700 in Figure 9-6). A rather large number of crossovers has to be attempted in order for the string to grow wider (frames 2000, 2200), i.e. into a strip of sp^2-hybridized carbon atoms (graphene) at frame 2200. Planar crossovers performed with parent structures that consist of graphene strips will likely lead to two flat strips, as shown in frame 2300. The two strips subsequently coalesce at an angle (frame 2700 in Figure 9-6), acquire more atoms and start curving onto a cylindrical surface due to the confining potential walls (frame 3300). Planar mating of parent structures such as that shown in frame 3300 results in closing the circumference, as shown in frame 3600; this closing occurs with defects along the cross-over planes, but such defects are systematically weeded out later on (see frames 4100 and 4300).

We have thus shown that a real space, variable-number algorithm can be used to retrieve simultaneously *the lowest-energy structure and the optimal number of atoms* of 1-D nanostructures subjected to desired conditions of radial confinement, starting from a single atom in the periodic unit cell. This algorithm is based on two-parent crossover operations and zero-penalty "mutations", the latter of which allowing for the algorithm to evolve even from a genetic pool made of identical structures. The structure of nanowires and nanotubes subjected to radial confinement were obtained for other materials as well (e.g. gold, Lennard-Jones atoms), which

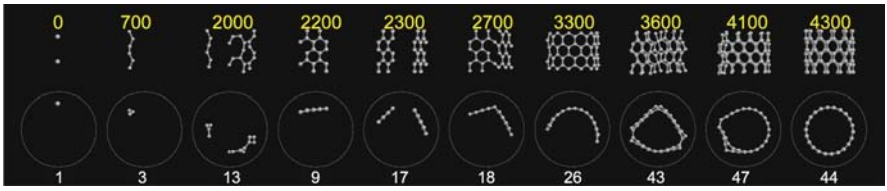


Figure 9-6. Evolution of lowest-energy carbon structure during a GA run performed with all crossover types. The system starts with only 1 atom for each member of the genetic pool, and evolves towards a defect-free CNT as the lowest-energy pool member. The outer confining wall is shown by the *white dash line*, and the "time" (i.e., the index of the crossover operation) is indicated by the *number* shown *atop* each frame. The number of atoms is shown at the *bottom* of the frames; for clarity, two periodic lengths along the CNT axis are displayed. Reproduced from [60], with permission from Taylor & Francis

suggests that the variable number genetic algorithm is a versatile tool for finding the atomic configuration of nanowires.

9.4. FUTURE DIRECTIONS

We have so far described, in certain detail, two methodologies (PTMC and GA) for the finding the structure of semiconductor surfaces. One of the two methods (GA) has also been designed for the study of nanowire structures. At this stage, the problem of pristine SiNWs structure that has triggered the development of GA for 1-dimensional system is still not solved, as we have chosen to study first a system (H-passivated Si [110] nanowire for which experimental observations [55] are available for comparison. The algorithm is to be developed over time, as one can add capabilities to it in order to study an increasing number of 1-dimensional material systems. An overview of the 1-D systems that can be tackled in the future is summarized in Figure 9-7, which shows the possible combinations between the materials targeted in this study, wire diameter regimes, and experimentally relevant surface terminations.

In terms of the semiconductor materials, silicon is perhaps the most important, and it is the material that drew our attention to the nanowire structure problem in the first place. Since, at its core, our methodology remains the same irrespective of the material chosen, we will consider not only Si but other materials as well. The binary materials (e.g., Si-Ge, In-P) in the chart above have been less scrutinized than silicon, and thus carry a tremendous potential for scientific novelty. The presence of a second atomic species in a sizeable proportion is bound to change the NW structure and will give insight into how the optical, electronic and mechanical properties can be tailored by changing the wire material and/or its composition.

The two regimes of wire diameters are delineated according to the current capabilities of the genetic algorithm, which turns out to be very efficient when dealing with less than 100 atoms per periodic cell along the wire axis. Defining the NW

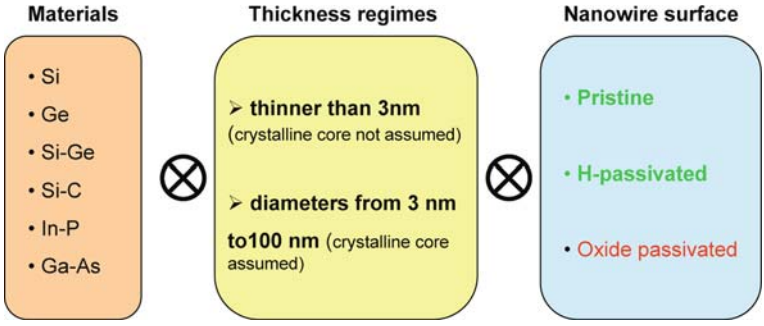


Figure 9-7. An overview of the nanowire semiconductor systems that might be tackled by global optimization methods. With the exception of the oxide-passivated wires (included here only for completness), the systems presented should be fairly amenable to study via genetic algorithms for diameters smaller than 10 nm. Reproduced from [54], with permission from Taylor & Francis

diameter as that of the smallest cylinder that includes all atoms, this number of atoms roughly corresponds to 3 nm diameter in the case of Si. For larger numbers of atoms per unit cell, the global optimization methods become much slower, as the difficulty of the structure determination problem increases exponentially; the methods should be further developed and refined to address, for example, a tripling of the number of atoms. Fortunately, in the regimes of diameters thicker than 3 nm, current experiments have been able to precisely identify a crystalline core with specific axis orientations [55, 56]. Therefore, for thick wires we will most likely not have to apply the GA, but instead study the structure and energetics of NWs starting from facet and facet-edge energies.

Most of the recent experiments show that the surface of the nanowires can fall into three main categories: clean [5], H-terminated [55] or oxide-terminated [62, 63]. Of these categories, the oxide surface is the least tractable with the currently available atomic interaction models. We cannot reasonably apply DFT methods either, because the structure of the oxide is mostly amorphous, thus hard to predict or assume. On the other hand, the clean and H-passivated NWs are very important for our fundamental understanding of the NWs structure and its effect on NW properties and applications.

Since in the nanometer-thin regime the structure of the NWs is crucial in determining novel phenomena, properties and applications, we believe that the impact of using global optimization methods for finding the structure of low-dimensional nanostructures should be significant, especially in the very foreseable future when the experimental community will achieve the stage at which it can routinely fabricate devices based on ultra-thin wires (smaller than 3 nm diameter). At that point, research in (e.g.) device conduction, optical phenomena, chemical sensing at the nanoscale, and nano-electromechanical properties will require immediate and detailed knowledge about the atomic structure as a starting point for any robust understanding of the operation of such devices. The methodologies that we have reviewed here have strong predictive capabilities, and therefore we hope that they will robustly complement the experimental techniques and provide the needed structure input for investigations of novel properties and functionalities of ultra-thin NWs.

ACKNOWLEDGMENTS

Ames Laboratory is operated for the United States Department of Energy by Iowa State University under Contract No. DE-AC02-07CH11358. CVC gratefully acknowledges support from the National Science Foundation under Grant. No. CMMI-0846858.

REFERENCES

1. S. Iijima, Nature **354**, 56 (1991).
2. D. Appell, Nature **419**, 553, (2002).

3. H. Kind, H.Q. Yan, B. Messer, M. Law, and P.D. Yang, Adv. Mater. **14**, 158, (2002).
4. Y. Cui and C.M. Lieber, Science **291**, 851, (2001).
5. B. Marsen and K. Sattler, Phys. Rev. B **60**, 11593, (1999).
6. M. Menon and E. Richter, Phys. Rev. Lett. **83**, 2973, (1999).
7. Y.F. Zhao and B.I. Yakobson, Phys. Rev. Lett. **91**, 035501 (2003).
8. R. Rurali and N. Lorente, Phys. Rev. Lett. **94**, 026805 (2005).
9. I. Ponomareva, M. Menon, D. Srivastava, and A.N. Andriotis, Phys. Rev. Lett. **95**, 265502, (2005).
10. R. Kagimura, R.W. Nunes, and H. Chacham, Phys. Rev. Lett. **95**, 115502 (2005).
11. M. Durandurdu, Physica Status Solidi B **243**, R7 (2006).
12. A.K. Singh, V. Kumar, R. Note, and Y. Kawazoe, Nano Lett. **5**, 2302 (2005).
13. W. Ranke and Y.R. Xing, Surf. Sci. **381**, 1 (1997).
14. T. Suzuki, H. Minoda, Y. Tanishiro, and K. Yagi, Surf. Sci. **358**, 522 (1996); *ibid.*, Surf. Sci. **348**, 335 (1996).
15. J. Liu, M. Takeguchi, H. Yasuda, and K. Furuya, J. Cryst. Growth **237–239**, 188 (2002).
16. S. Joeng, H. Jeong, S. Cho, and J.M. Seo, Surf. Sci. **557**, 183 (2004).
17. A.A. Baski, S.C. Erwin, and L.J. Whitman, Science **269**, 1556 (1995).
18. J. Dabrowski, H. J. Müssig, and G. Wolff, Phys. Rev. Lett. **73**, 1660 (1994).
19. G.D. Lee and E. Yoon, Phys. Rev. B **68**, 113304 (2003).
20. Y.W. Mo, D.E. Savage, B.S. Swartzentruber, and M.G. Lagally, Phys. Rev. Lett. **65**, 1020 (1990).
21. R.G. Zhao, Z. Gai, W. Li, J. Jiang, Y. Fujikawa, T. Sakurai, and W.S. Yang, Surf. Sci. **517**, 98 (2002).
22. Y. Fujikawa, K. Akiyama, T. Nagao, T. Sakurai, M.G. Lagally, T. Hashimoto, Y. Morikawa, and K. Terakura, Phys. Rev. Lett. **88**,176101 (2002).
23. P. Raiteri, D.B. Migas, L. Miglio, A. Rastelli, and H. von Kánel, Phys. Rev. Lett. **88**, 256103 (2002).
24. V.B. Shenoy, C.V. Ciobanu, and L.B. Freund, Appl. Phys. Lett. **81**, 364 (2002).
25. C.V. Ciobanu, V.B. Shenoy, C.Z. Wang, and K.M. Ho, Surf. Sci. **544**, L715 (2003).
26. D.P. Woodruff, Surf. Sci. **500**, 147 (2002).
27. K.E. Khor and S. Das Sarma, J. Vac. Sci. Technol. B **15**, 1051 (1997) 1051.
28. F.H. Stillinger and T.A. Weber, Phys. Rev. A **28**, 2408 (1983).
29. F.H. Stillinger, Phys. Rev. E **59**, 48 (1999).
30. D.J. Wales and J.P.K. Doye, J. Phys. Chem. A **101**, 5111 (1997).
31. A.F. Voter, Phys. Rev. B **57**, R13985 (1998).
32. M.R. Sørensen and A.F. Voter, J. Chem. Phys. **112**, 9599 (2000).
33. C.J. Geyer and E.A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995).
34. K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
35. C.V. Ciobanu and C. Predescu, Phys. Rev. B **70**, 085321 (2004).
36. D.M. Deaven and K.M. Ho, Phys. Rev. Lett. **75**, 288 (1995).
37. K.M. Ho, A.A. Shvartsburg, B.C. Pan, Z.Y. Lu, C.Z. Wang, J. Wacker, J.L. Fye, and M.F. Jarrold, Nature **392**, 582 (1998).
38. F.C. Chuang, C.V. Ciobanu, V.B. Shenoy, C.Z. Wang, and K.M. Ho, Surf. Sci. **573**, L375 (2004).
39. F.C. Chuang, C.V. Ciobanu, C. Predescu, C.Z. Wang, and K.M. Ho, Surf. Sci. **578**, 183 (2005).
40. F.C. Chuang, C.V. Ciobanu, C.Z. Wang, and K.M. Ho, J. Appl. Phys. **98**, 073507 (2005).
41. C.V. Ciobanu, F.C. Chuang, and D. Lytle, Appl. Phys. Lett. **91**, 171909 (2007).
42. T.J. Lenosky, B. Sadigh, E. Alonso, V.V. Bulatov, T. Diaz de la Rubia, J. Kim, A.F. Voter, and J.D. Kress, Model. Simul. Mater. Sci. Eng. **8**, 825 (2000).
43. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.M. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
44. U.H.E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
45. Y. Sugita, A. Kitao, and Y. Okamoto, J. Chem. Phys. **113**, 6042 (2000).
46. C. Predescu, M. Predescu, and C.V. Ciobanu, J. Chem. Phys. **120**, 4119 (2004).

47. S. Baroni, A. Dal Corso, S. de Gironcoli, and P. Giannozzi, http://www.pwscf.org.
48. F.H. Stillinger and T.A. Weber, Phys. Rev. B **31**, 5262 (1985).
49. J. Tersoff, Phys. Rev. B **38**, 9902 (1988);*ibid* Phys. Rev. B **37**, 6991 (1988).
50. S.C. Erwin, A.A. Baski, and L.J. Whitman, Phys. Rev. Lett. **77**, 687 (1996).
51. L. Hui, B.L. Wang, J.L. Wang, and G.H. Wang, J. Chem. Phys. **121**, 8990 (2004).
52. B.L. Wang, G.H. Wang, and J.J. Zhao, Phys. Rev. B **65**, 235406, (2002).
53. B.L. Wang, S.Y. Yin, G.H. Wang, A. Buldum, and J.J. Zhao, Phys. Rev. Lett. **86**, 2046(2001).
54. C.V. Ciobanu, C.Z Wang, and K.M. Ho, Mater. Manuf. Process. **24**, 109, (2009).
55. D.D.D. Ma, C.S. Lee, F.C.K. Au, S.Y. Tong, and S.T. Lee, Science **299**, 1874, (2003).
56. Y. Wu, Y. Cui, L. Huynh, C.J. Barrelet, D.C. Bell, and C.M. Lieber, Nano Lett. **4**, 433, (2004).
57. T.L. Chan, C.V. Ciobanu, F.C. Chuang, N. Lu, C.Z. Wang, and K.M. Ho, Nano Lett. **6**, 277 (2006).
58. U. Hansen and P. Vogl, Phys. Rev. B **57**, 13295 (1998).
59. VIENNA ab initio simulation package, Technische Universitat Wien, 1999; G. Kresse and J. Hafner, Phys. Rev. B **47**, R558 (1993); G. Kresse and J. Furthmuller, Phys. Rev. B **54**, 11169 (1996).
60. T.E.B. Davies, D.P. Mehta, J.L. Rodriguez-Lopez, G.H. Gilmer, and C.V. Ciobanu, Mater. Manuf. Process. **24**, 265, (2009).
61. N.L. Abraham and M.I.J. Probert, Phys. Rev. B **73**, 224104, (2006).
62. Y. Cui, L.J. Gudiksen, M.S. Wang, and C.M. Lieber, Appl. Phys. Lett. **78**, 2214, (2001).
63. N. Wang, Y.H. Tang, Y.F. Zhang, C.S. Lee, I. Bello, and S. T. Lee, Chem. Phys. Lett. **299**, 237, (1999).

CHAPTER 10

ATOMIC-SCALE SIMULATIONS OF THE MECHANICAL BEHAVIOR OF CARBON NANOTUBE SYSTEMS

BYEONG-WOO JEONG[1,2] AND SUSAN B. SINNOTT[2]

[1]*School of Mechanical Engineering, Yeungjin College, Taegu, 702-721, South Korea, e-mail: bwoojeong@gmail.com*

[2]*Department of Materials Science and Engineering, University of Florida, Gainesville, FL 32611-6400, USA,e-mail: ssinn@mse.ufl.edu*

Abstract: Among the objects of nanomechanics studies, carbon nanotubes have attracted special interest due to their unique properties and potential use in a wide range of applications, including nanometer-scale devices and composite materials. In these applications the mechanical responses of the nanotubes can significantly affect the performance of the devices and materials. Therefore, characterizing and understanding their mechanical responses is necessary in order to optimize their utilization in these applications. Computational simulations are uniquely able to provide insights that are challenging to obtain experimentally. Molecular dynamics simulations in particular are popular for the examination of the mechanical responses of nanotubes.

This chapter provides a review of the background of molecular dynamics simulation methods, their role in the study of the nanomechanical responses of carbon nanotubes, and their important contributions to this emerging research field. Illustrative examples are presented that illustrate how these approaches are providing new and exciting insights into nanomechanical properties as elastic modulus or stiffness, fracture, and bucklng. Furthermore, the simulations indicate that filling or functionalization, combined loads, and external gases influence these properties. Thus, molecular dynamics simulation methods are revolutionizing our understanding of the mechanical behavior of nanotube systems at the most fundamental atomic level

Keywords: Molecular dynamics simulation, Carbon nanotube, Nanomechanics

10.1. INTRODUCTION

Nanomechanics is an emerging research area that deals with the mechanical properties and behavior of nanometer-scale structures and materials [1]. Among the numerous objects of nanomechanics studies, carbon nanotubes (CNTs) have attracted special interest due to their unique properties and potential use in a wide

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 255–295.
DOI 10.1007/978-1-4020-9785-0_10,

range of applications. In fact, CNTs have the unique properties including high stiffness, high covalent bond strength, large elastic instability, low density, tubular shape, and large aspect ratio. They have consequently been proposed for use as key elements in applications such as nanometer scale devices [2–14] and composite materials [15]. In these CNT-based applications, the mechanical properties and behavior of the CNTs are key and can significantly influence the performance of the resulting devices or materials. Therefore, characterizing and understanding the mechanical responses of CNTs is important to optimize the performance of CNT-based applications.

Computational simulations are uniquely able to address problems that cannot be readily addressed analytically or experimentally. Thus, they have been commonly used to evaluate and predict the mechanical properties and behavior of CNTs as it is non-trivial to manipulate CNTs experimentally due to their small size [16–29]. A variety of computational simulation methods have been used to investigate the mechanical responses of CNTs, including classical molecular dynamics (MD) simulations with empirical potentials, Monte Carlo simulations, tight-binding calculations and simulations, density functional theory calculations and simulations, and well-defined continuum mechanics methods. Among these various methods, classical MD simulations have been one of the most popular tools for the examination of the mechanical responses of CNTs. For example, they have provided insight into their elastic [30–37] and plastic [38–46] behavior, their interactive behavior with other materials [47–51], their dynamic responses [52], among others [53–58]. MD simulations are similar to real experiments in many respects in that events are predicted in real time and the approach is able to consider the qualitative dynamical behavior of relatively large numbers of atoms.

In MD simulations, atomic trajectories are calculated by numerically integrating coupled classical equations of motion. Interatomic forces that enter these equations are typically calculated either from total energy methods that include electronic degrees of freedom, or from simplified mathematical expressions that give the potential energy as a function of interatomic displacements. MD simulations can thus be considered numerical experiments that provide a link between analytic models and experiments. The another main strength of MD simulations is that they can reveal unanticipated phenomena or unexpected mechanisms for well-known observations. The next section contains a review of MD simulations, including the approximations that are inherent in their application to the study of the nanomechanical properties of CNTs. The subsequent section discusses some of the important insights and findings that have been obtained from MD simulations of the mechanical behavior of nanotubes, such as determining their elastic moduli or stiffness, their fracture and bucklng behaviors, the influence of filling or chemical functionalization, the effect of combined loads, and the consequences of interactions with external gases. These studies have not only provided insights into the physical phenomena, they have also revealed a wealth of atomic-scale phenomena that occur during the mechanical behavior of various nanotube systems that was not previously known.

10.2. COMPUTATIONAL DETAILS

The MD simulation method was first introduced by Alder and Wainwright in the late 1950's to study the interactions of hard spheres [59, 60]. MD simulations have been improved with many algorithms and tools since then and can now be used to investigate various system states, such as gases, liquids, surfaces, bulk defects, fracture phenomena, friction, and biomaterial interactions. MD simulations are the process that generates atomic trajectories of a system of N particles by direct numerical integration of Newton's equations of motion with appropriate specification of an interatomic potential and suitable initial and boundary conditions. Unlike Monte-Carlo (MC) simulations, MD simulations can provide detailed information about how the system evolves during a given process. Because of this, the results of MD simulations can be directly compared with real experiments in most cases.

MD simulations are straightforward to describe: given a set of initial conditions and a way of mathematically modeling interatomic forces, Newton's (or equivalent) classical equation of motion is numerically integrated:

$$F = ma, \tag{10-1a}$$

$$-\nabla E = m(\partial^2 r/\partial t^2), \tag{10-1b}$$

where F is the force on each atom, m is the atomic mass, a is the atomic acceleration, E is the potential energy felt by each atom, r is the atomic position, and t is time. The forces acting on any given atom are calculated, and then the atoms move a short increment ∂t (called a time step) forward in time in response to these applied forces. This is accompanied by a change in atomic positions, velocities, and accelerations. The process is then repeated for some specified number of time steps.

The output of these simulations includes new atomic positions, velocities, and forces that allow additional quantities such as temperature and pressure to be determined. As the size of the system increases, it is useful to render the atomic positions in animations that reveal the responses of the system in a qualitative manner. Quantitative data can be obtained by analyzing the numerical output directly. The following sub-sections review the way in which energies and forces are calculated in MD simulations and the important approximations that are used to realistically model the mechanical behavior of nanometer-scale structures such as CNTs. The reader is referred to additional sources [61–64] for a more comprehensive overview of MD simulations and the interatomic potentials.

10.2.1. Interatomic Potentials

There are several different approaches by which interatomic potentials are determined in MD simulations. The most theoretically accurate methods are those such as *ab initio* or first principles. These approaches are derived from quantum mechanical theory and are generally both the most accurate and the most computationally

intensive. They are therefore limited to a small number of atoms, which has limited their use in the study of mechanical behavior. Alternatively, empirical methods contain parameters that are determined by fitting to experimental data or the results of ab initio calculations. These techniques can usually be used to correctly describe qualitative trends and are often the only choice available for modeling relatively large systems. Empirical methods have therefore been widely used in studies of the mechanical behavior of relatively large systems and materials.

Empirical methods simplify the modeling of materials by treating the atoms as spheres. Therefore, electrons are not treated explicitly in this approach, although it is understood that the interatomic interactions are ultimately dependent on them. The spheres interact with each other via repulsive and attractive functional terms that generally depend on interatomic distances and/or angles and contain adjustable parameters that are fit to *ab initio* results and/or experimental data. The main strength of empirical potentials is their computational speed and thus they can be used to simulate relatively large-size systems. There are several important and common general classes of empirical methods used for calculating interatomic potentials, including the bond-order potential (BOP) [65], the reactive empirical bond-order (REBO) potential [66–69], the Stillinger-Weber potential [70], the embedded atom method (EAM) [71, 72] and modified embedded atom method (MEAM) [73, 74] approach, and Coulomb or multipole interaction potentials [75, 76]. Out of all of these potentials, only the REBO potential is reviewed here as it is most applicable to the study of CNTs.

The reactive empirical bond-order potential was initially developed and parameterized by Tersoff [66, 67] to model materials such as carbon and silicon. It was based on the formalism of Abell [65] and included a many-body interactions. While the Tersoff potential can describe the carbon-carbon single, double, and triple bond lengths and energies for hydrocarbons, solid graphite, and diamond, it cannot describe bonding situations intermediate between single and double bonds. To correct this and the non-physical overbinding of radicals, Brenner [68] developed an improved form of Tersoff-type potential for hydrocarbons. There was, however, a limit to model processes involving energetic atomic collisions in the Tersoff potential because the Morse-type functions for pair interactions go to finite values as the distance between atoms decreases. Accordingly, Brenner et al. [69] modified the expressions for interatomic interactions and expanded the fitting database. This is so-called second generation REBO potential that yields more accurate bond lengths, energies, and force constants for hydrocarbons. It is generally established that the REBO potential can predict realistic physical properties and behavior for CNTs [33–35, 39, 40, 42, 52–54].

The expression of the REBO potential used to calculate the binding energy (E_b) between atoms i and j is:

$$E_b = \sum_i \sum_{j>i} \left[V_R\left(r_{ij}\right) - b_{ij} V_A\left(r_{ij}\right) \right] \tag{10-2}$$

where $V_R(r_{ij})$ and $V_A(r_{ij})$ are repulsive and attractive pairwise potentials determined by the atom types i and j, and that only depend on the distance r_{ij} between

the two atoms. The potential is short-ranged and only considers nearest neighbor bonds. To model long-range nonbonded interactions, the REBO potential is combined with pair-wise potentials either directly through splines [37] or indirectly with more sophisticated functions [77]. The terms, $V_{\mathrm{R}}(r_{ij})$ and $V_{\mathrm{A}}(r_{ij})$ of Eq. (10-2) are given as,

$$V_{\mathrm{R}}(r_{ij}) = f_{\mathrm{c}}(r_{ij})\left[1 + \frac{Q}{r_{ij}}\right] \cdot A \cdot e^{-\alpha \cdot r_{ij}} \tag{10-3a}$$

$$V_{\mathrm{A}}(r_{ij}) = f_{\mathrm{c}}(r_{ij}) \sum_{n=1}^{3} B_n \cdot e^{-\beta_n \cdot r_{ij}} \tag{10-3b}$$

where *A, B, Q,* α, and β are parameters determined by atom type. The function $f_{\mathrm{c}}(r_{ij})$ is a cutoff function that limits the range of the covalent interactions to insure that the interactions include nearest neighbors only.

The many-body empirical bond-order term, b_{ij}, between atom i and j in Eq. (10-2) represents the many-body feature of the Tersoff type potential. It includes various chemical effects such as coordination numbers, bond angles, torsion angles, and conjugation effects and it depends on the local atomic environment in which a particular bond is located. Therefore, the REBO potential can describe covalent bond formation and breakage associated with atomic hybridization alteration by weighing the bond strength. This term is most essential for treating chemical reactions, in which the bonding of carbon atoms changes, and is written as a sum of terms:

$$b_{ij} = \frac{1}{2}\left[b_{ij}^{\sigma-\pi} + b_{ji}^{\sigma-\pi}\right] + b_{ij}^{\pi} \tag{10-4}$$

where $b_{ij}^{\sigma-\pi}$ and $b_{ji}^{\sigma-\pi}$ represent the local coordination and bond angles for atoms i and j. The term b_{ij}^{π} is further written as a sum of two terms:

$$b_{ij}^{\pi} = \Pi_{ij}^{\mathrm{RC}} + \Pi_{ij}^{\mathrm{DH}} \tag{10-5}$$

where Π_{ij}^{RC} term describes conjugated system and radical character between atoms i and j, and Π_{ij}^{DH} term depends on the dihedral angle for carbon-carbon bonds which considers the torsion effect in the molecule.

Long-range van der Waals or related forces are typically modeled with pairwise additive potentials. A widely used approximation is the Lennard-Jones (LJ) potential [63, 78], which has the following functional form:

$$V_{\mathrm{LJ}}(r_{ij}) = 4 \cdot \varepsilon \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{10-6}$$

where σ and ε are the L-J parameters for each particular type of atom being modeled and r_{ij} is the interatomic distance between atoms i and j.

In the MD simulations reviewed here, the forces on the atoms are calculated using methods that vary with interatomic distance. The REBO potential for the short-range covalent interactions is coupled with Lennard-Jones potential for the long-range van der Waals interactions in the MD simulations

10.2.2. Important Approximations

There are several important approximations that are used to realistically model the systems in MD simulations. They are discussed as follows.

10.2.2.1. Periodic Boundary Conditions

The purpose of molecular simulations is to model the macroscopic sample at the atomic scale and provide information that is not easily obtainable from experiments. Unfortunately, due to the computational limitations of present-day computers, the number of atoms that can be conveniently handled ranges from a few hundred to a few billion. In order to model a macroscopic system in terms of a finite simulation system of N particles, periodic boundary conditions (PBC) are employed. The concept of periodic boundary conditions is illustrated in Figure 10-1, where the simulation system of N particles is treated as a basic unit and is replicated throughout space. Therefore, the simulation unit is essentially embedded in an infinite array of units. In the figure, shaded box (primary cell) represents the system simulated.

The periodic boundary conditions do not need to be applied if the systems simulated are small and there is no need to mimic larger systems [30, 33, 36]. However, if the systems simulated are large, periodic boundary conditions allow MD

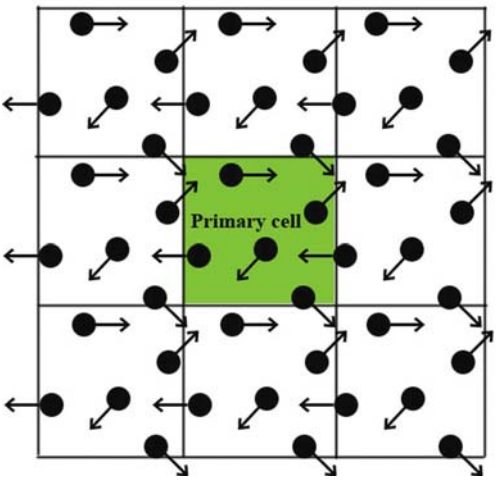


Figure 10-1. Schematic representation of periodic boundary condition (PBC). Primary cell (*shaded box*) represents the system simulated

simulations with a relatively small number of atoms to accurately mimic much larger systems [53, 55, 56].

10.2.2.2. Temperature Control

In basic MD simulations, the system has a constant value of particles (N), volume (V), and energy (E) because the total energy should be conserved in any closed system that follows Newton's equation of motion. Therefore, in this basic MD simulation, the system properties are measured in the microcanonical (constant NVE) ensemble [61]. However, in real cases, it may be necessary to perform simulations at constant system temperature (T) because most real experiments are carried out at constant temperatures. In this case, the properties of the system is measured in the canonical (constant NVT) ensemble [61]. In order to perform MD simulations at constant temperature, the simulation system is brought into contact with a thermostat. A schematic diagram of a typical thermostat application is shown in Figure 10-2. In the figure, atoms in the active region follow Newton's equation of motion, and thus the properties of the systems of interest are unaffected by the thermostat.

Several methods, such as velocity-rescaling, Andersen, Nosé-Hoover, and Langevin thermostats, are used to control system temperature in MD simulations [61, 63]. In particular, the velocity-rescaling thermostat is widely adopted to investigate the mechanical behavior of CNTs as it has been shown to have negligible effects on such behavior [30, 33, 39]. In this thermostat, the velocity is rescaled to keep the kinetic energy constant and thus to control the system temperature as follows:

$$v_{\text{new}} = v_{\text{old}} \times \sqrt{\frac{T_0}{T_{\text{ins}}}} \tag{10-7}$$

where v_{new} is the rescaled velocity, v_{old} is the velocity prior to rescaling, T_0 is the reference temperature of the thermostat, and T_{ins} is the instantaneous temperature at

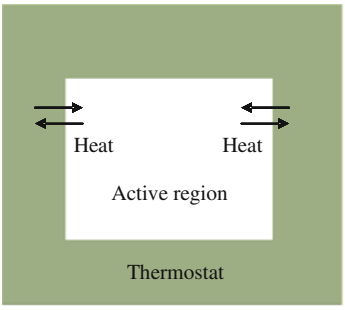


Figure 10-2. Schematic diagram of typical thermostat applications

any time. The velocity-rescaling thermostat is a straightforward approach that has been used since the early days of MD simulations.

Another popular thermostat is the Langevin thermostat in which the atoms in the thermostat region follows the Langevin equation of motion instead of Newton's equations of motion [61]. The Langevin equation of motion is expressed as:

$$m \cdot a = -\xi \cdot v + f(r) + f' \tag{10-8}$$

where m is the mass of the particle, a is the acceleration, $f(r)$ is the force obtained from interatomic potential, v is the velocity of the particle, ξ is a friction constant, and f' is the random force. The effect of friction force $\xi \cdot v$ decreases the temperature of the system because ξ is positive. The random force is determined from a Gaussian distribution, and its variances are the desired temperature and time step. Therefore, the random force is balanced with the frictional force and maintains the system temperature in a desired manner. It should be noted that the random forces are uncoupled from those at previous steps, which is denoted by the delta function. Additionally, the width of the Gaussian distribution from which the random force is obtained varies with temperature. Thus, the Langevin approach does not require any feedback from the current temperature of the system.

Andersen [79] proposed that the pressure and/or the temperature should be held constant in some MD simulations to better approximate experimental conditions. In Anderson's method, a particle is randomly chosen and its velocity is extracted from the Maxwell distribution. This simple approach worked but was known to have poor efficiency and discontinuous trajectories [61]. Anderson's approach, however, inspired Nosé [80] who proposed an extended system method for the canonical ensemble. He introduced a time scale variable s and its conjugate momentum p_s, which describe the coupling of system to the thermostat. An additional parameter Q can be regarded as a thermostat mass. Hoover [81] simplified Nosé's approach by eliminating the time scaling factor s and introducing a thermodynamic friction coefficient ζ. Hoover's modified expression of Nosé's approach is known as the Nosé-Hoover thermostat.

As indicated in this discussion, thermostat methods have their own strengths and weaknesses that should be well-understood prior to their use and that influence the interpretation of results. Therefore, the choice of thermostat method and the fraction of thermostat atoms are important to correctly examine the behavior of materials [30].

10.2.2.3. Predictor-Corrector Algorithm

The predictor-corrector algorithm is one of the most widely used algorithms in MD simulations. The basic idea of the predictor-corrector is that the position, velocity, and acceleration of each particle at time $t+\Delta t$ are predicted first by Taylor expansion, and then the predicted values are corrected by the interatomic forces calculated from the interatomic potentials and Newton's equation of motion. The form of the predictor is expressed as [61]:

$$\begin{cases} r_{\mathrm{p}}(t+\Delta t) = r(t) + v(t)\cdot\Delta t + \frac{1}{2}a(t)\cdot\Delta t^2 + \frac{1}{6}b(t)\cdot\Delta t^3 \\ v_{\mathrm{p}}(t+\Delta t) = v(t) + a(t)\cdot\Delta t + \frac{1}{2}b(t)\cdot\Delta t^2 \\ a_{\mathrm{p}}(t+\Delta t) = a(t) + b(t)\cdot\Delta t \\ b_{\mathrm{p}}(t+\Delta t) = b(t) \end{cases} \tag{10-9}$$

where r_{p}, v_{p}, a_{p} and b_{p} are the predicted position, velocity, acceleration and third derivative of each atom at $t + \Delta t$. Then the interatomic forces are calculated based on the predicted position of each atom and the corrected accelerations, a_{c} $(t+\Delta t)$. Usually there will be a discrepancy between a_{p} $(t+\Delta t)$ and a_{c} $(t+\Delta t)$. The adjustment parameter, Δa $(t+\Delta t)$, is used in the corrector to bring the predicted values into agreement with the calculated values, and is defined as $\left[a_{\mathrm{c}}(t+\Delta t) - a_{\mathrm{p}}(t+\Delta t)\right]$. Then, the position and other derivatives can be corrected using the following equations.

$$\begin{cases} r_{\mathrm{c}}(t+\Delta t) = r_{\mathrm{p}}(t+\Delta t) + \frac{1}{6}\Delta a(t+\Delta t) \\ v_{\mathrm{c}}(t+\Delta t) = v_{\mathrm{p}}(t+\Delta t) + \frac{5}{6}\Delta a(t+\Delta t) \\ a_{\mathrm{c}}(t+\Delta t) = a_{\mathrm{p}}(t+\Delta t) + \Delta a(t+\Delta t) \\ b_{\mathrm{c}}(t+\Delta t) = b_{\mathrm{p}}(t+\Delta t) + \frac{1}{3}\Delta a(t+\Delta t) \end{cases} \tag{10-10}$$

These corrected values are used to predict the positions and first n derivatives at the next step in the trajectory, and then the same procedure is repeated until the end of the trajectory is reached. Equations (10-9) and (10-10) as written are termed a third order predictor-corrector algorithm within the Nordseck version [61].

Although higher order predictor-corrector algorithms, that take higher order derivatives of position with respect to time, can give more accurate results, the accuracy of the trajectory is more sensitive to the length of the time step than to the order of the algorithm. The time step should be, therefore, determined to obtain an optimal compromise between accuracy and simulation time. In highly dynamic systems, short time steps should be used.

10.2.2.4. Simulation Methods for Mechanical Behavior

In an early study of the mechanical behavior of CNTs, Yakobson et al. [1] deformed CNTs and examined their mechanical responses using the displacement control method. In this approach, displacement is given to the CNTs and then the resulting force is calculated from the potential energy in MD simulations. Thus, this method is similar to the displacement control method that is used for the property test of macroscopic materials.

On the other hand, in the load control method, the external loads are applied to CNTs and then displacements are calculated in MD simulations [33–35, 39, 40]. Therefore, this approach is similar to the load control method that is used for the testing of macroscopic materials. The advantage of this method is that definite external loads can be applied within the simulation systems. In this approach, the external load on each atom is added to the force on each atom calculated by the potential.

Here, the total external load is the sum of the external load on each atom. The basic concept of this method is given by Eq. (10-11).

$$F_i = -\frac{\partial U}{\partial r_i} + F_i^{\text{external}} \tag{10-11a}$$

$$F_{\text{total}}^{\text{external}} = \sum_{1}^{n} F_i^{\text{external}} \tag{10-11b}$$

where, F_i is the force on atom i, F_i^{external} is the external force on atom i, U is the interatomic potential, $F_{\text{total}}^{\text{external}}$ is the total external force applied to the CNTs, and n is the number of atoms loaded externally. Specially, in the case of torsional moment, the external force on each atom is applied as the tangential force on the atoms.

10.3. MECHANICAL BEHAVIOR OF NANOTUBES

In the most general terms, failure refers to any actions leading to an inability of a part to function in the intended manner. It follows that permanent deformation (yielding) or fracture may be regarded as modes of failure. Another way in which a part may fail is through instability by undergoing large displacements when the applied load reaches the buckling value. Here, while the permanent deformation or fracture indicates the elastic-plastic yield transition of parts, buckling instability is an effect of overall geometry rather than only material instability. Therefore, the beginnings of failure occur prior to the onset of any high levels of stress and the buckled systems are totally elastic. At macroscopic length scales, the failures of columns or thin walled structures under compressive, bending, or torsional load are governed by buckling instability and thus the buckling rather than strength considerations dictates the performance of these structures.

Although CNTs are nanometer-scale structures, their overall shape resembles macroscopic thin-walled hollow columns or tubes. Accordingly, it has been shown that the investigation of mechanical behavior of CNTs under compression, bending, or torsion should also be considered from the point of view of buckling instability [1]. While the CNTs have their highest rigidity along the direction of the nanotube axis due to the sp^2 bonding between the carbon atoms, they are much more compliant in their radial direction [1, 33, 36]. Therefore, the CNTs are readily collapsed in their radial direction without breaking any in-plane σ-bonding [1, 33, 36]. This characterizes the buckling instability which occurs in the CNTs under compressive, bending, or torsional load. However, the CNTs under tensile load exhibit the elastic-plastic yield transition or fracture behavior [38, 41, 45].

Computational simulations have been widely used to evaluate and predict the properties and behavior of CNTs as is still relatively challenging to do so experimentally due to the nanometer-scale size of the nanotubes [16–29]. In particular, MD simulations are similar to real experiments in many respects in that events are predicted in real time, and are suitable for examining the qualitative dynamical behavior of relatively large numbers of atoms. Therefore, MD simulation has

been the most commonly used to investigate the mechanical responses of CNTs using appropriate empirical interatomic potentials [30–58]. The rest of this section discusses some of the important insights and findings that have been obtained from MD simulations of the mechanical responses of the CNTs such as their tensile, compressive, bending, and torsional responses to loading. Importantly, these studies have revealed atomic-scale details of the mechanical responses of the nanotube systems.

10.3.1. Tensile Behavior

It has been established that nanotubes are more compliant in their radial direction than in their axial direction, and, in fact, have high rigidity in their axial direction. Consequently, the tensile behavior of the nanotubes is expected to be different from their behavior in response to compression, bending, or torsion.

10.3.1.1. Young's Modulus

Analysis of the elastic moduli of CNTs is made by assuming they are elastic beams. It is therefore necessary to determine their volume. Equation (10-12) provides the definition of tensile elastic modulus (or Young's modulus) based on continuum mechanics theory [82, 83]. When applied to CNTs, the Young's moduli of the CNTs are calculated directly from their strain energy and volume as follows:

$$Y = \frac{2U}{V\varepsilon^2} = \frac{2U}{AL\varepsilon^2} \tag{10-12}$$

where V is the volume of the CNT, A is the cross-sectional area, L is length, U is the strain energy, and ε is the strain. In Eq. (10-12), the sectional area of CNTs can be calculated from the effective wall thickness of CNTs which has been commonly taken to be 0.34 nm, which is the interplanar distance between graphene layers [84].

In most studies, the Young's moduli of CNTs have been predicted to be about 0.9–1.2 TPa by assuming 0.34 nm as the effective wall thickness of CNTs. For example, WenXing et al. [85] calculated the Young's modulus of single-walled carbon nanotubes (SWNTs) from MD simulations, and reported that the value is about 0.95 TPa and weakly affected by the tube chirality and radius. Using both empirical potentials and first principle methods, Robertson et al. [86] examined the strain energies and the elastic properties of nanotubes having radii less than 0.9 nm and predicted an elastic modulus of about 1.06 TPa. Lu [84] calculated a value of about 0.97 TPa for the moduli of various CNTs and predicted that the variation with CNT diameter is insignificant. Hernandez et al. [87] predicted a value of 1.2 TPa using tight-binding method. Experimental investigations of the elastic moduli of CNTs are somewhat limited due to the difficulties involved in the production of defect-free CNTs and experimental techniques, and thus there are many fewer measurements relative to calculated values. The first experimental determination of elastic modulus was done by Treacy et al. [88] who measured the thermal vibrations of free-standing multi-walled carbon nanotubes (MWNTs) inside a

transmission electron microscope. An average elastic modulus of 1.25 TPa was reported. Yu et al. [89] performed experimental investigation on tension of SWNTs in a scanning electron microscope and then obtained an average value of 1.0 TPa.

However, Yakobson et al. [45] calculated the value of about 5.5 TPa that is not realistic and much higher than values mentioned above. This is because a value of 0.066 nm (the thickness of an individual atom of carbon) used as effective wall thickness was much smaller than 0.34 nm used in above studies. Consequently, when the Young's modulus of CNTs is calculated based on continuum mechanics model, the definition of their wall thickness is important and the examples discussed above demonstrate that the value of 0.34 nm is reasonable assumption as the effective wall thickness of CNTs.

10.3.1.2. Fracture or Plastic Behavior

The failures of CNTs under compressive, torsional, or bending load are governed by buckling instability and thus buckling rather than strength considerations dictates their performance. However, the failure of the CNTs under tensile load should be dominated by the elastic-plastic yield transition or fracture behavior. In particular, perfect nanotubes (defect free nanotubes) undergo brittle fracture, while nanotubes with vacancies exhibit plastic necking and thinning [41]. This fracture [39, 42] or plastic behavior [41, 45] of CNTs has been successfully predicted from MD simulations.

Some researchers have examined the fracture of CNTs with MD simulations, usually with the second-generation REBO potential. However, it has been previously demonstrated that the cutoff functions of this potential overestimate the force needed to break a carbon-carbon covalent bond [21, 29, 38, 42]. This overestimation is due to the form of the cutoff functions in the potential, which artificially raises the bond force for distances between 1.7 and 2.0 Å [21, 29, 38, 42]. In order to prevent this overestimation, researchers have modified these cutoff functions. Jeong et al. [39] and Sammalkorpi et al. [42] increased the onset of the covalent interaction cutoff distance from 1.7 to 1.95 Å. This modification does not have any adverse effects on the potential under tensile load. Figure 10-3 shows the typical tension-tensile strain curves that result [39]. The figure indicates that the tensile fracture loads and the influence of filling materials are overestimated using the unmodified cutoff functions. The overestimation as a result of filling is due to a Poisson ratio effect that contracts the CNTs in the radial direction [29]. In the filled CNTs under tension, the radial-direction contraction decreases, leading to increased van der Waals interaction forces between the CNTs and filling materials that influence the progress of tensile strain, in agreement with the results of previous work [45, 46]. This is especially apparent as the radial-direction contraction increases above 30% tensile strain. In contrast, when modified cutoff functions are used, the tensile fracture loads are substantially decreased and relatively unaffected by CNT filling.

In most MD simulations, the tensile strength is determined to be about 100 GPa [38, 39, 42], but this value is higher than the tensile strength of about 50 GPa obtained from experiments [89, 90]. This discrepancy is thought to be caused mainly

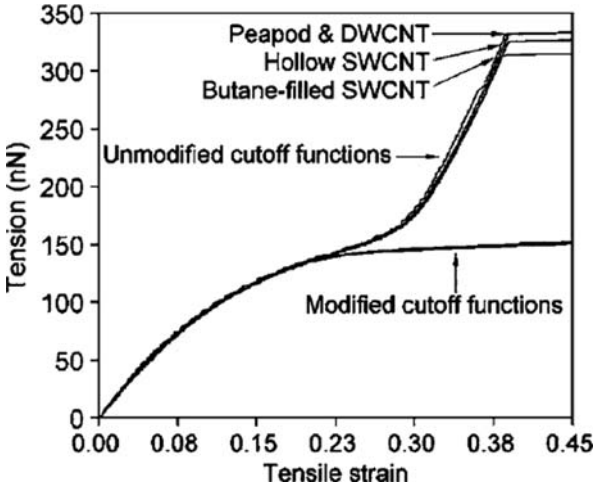


Figure 10-3. Tension-tensile strain curves of hollow and filled single-walled carbon nanotubes (SWCNTs) under tension using unmodified or modified cutoff functions in the second-generation REBO potential. DWCNT is a double-walled carbon nanotube. After [39] with permission of AIP (2007)

by the wall defects and other imperfections in the experimental CNT samples that are not present in the CNTs examined computationally.

10.3.1.3. Effect of Filling, Functionalization, and Temperature

There has also been considerable interest in filling CNTs with various materials [91] and in chemically modifying CNTs [92], which can affect their mechanical response. Jeong et al. [39] investigated the effect of functionalization and filling on the tensile behavior at various temperatures using MD simulations. The CNTs were filled with n-butane or C_{60} of which the densities are 0.53 and 1.21 g/cm^3, respectively. Their simulations indicated that filling CNTs does not affect the tensile fracture load as well as tensile elastic modulus due to weak van der Waals interactions between the filling materials and CNT (Figure 10-3). Here, the effect of the tensile elastic modulus is in agreement with the results of previous studies of compression of CNTs filled with the same materials [36, 51]. Thus, filling CNTs does not alter their axial elastic modulus.

Since the chemical modification of CNT walls and temperature have been shown to influence the mechanical properties of CNTs [36, 44, 49], the effect on the tensile behavior of functionalization and temperature was considered by Jeong et al. [39] who also used MD simulations. Figure 10-4 [39] illustrates how functionalization by H_2C–C groups decreases the tensile fracture load and its decrement depends on the densities of the functional groups, which are slightly lower than those reported experimentally [92]. Specifically, the rates of decrement are about 1.2 and 3.0%, respectively. These values are lower than the 12.5% decrease predicted for compression of CNTs functionalized with the same groups, but in ordered configurations and

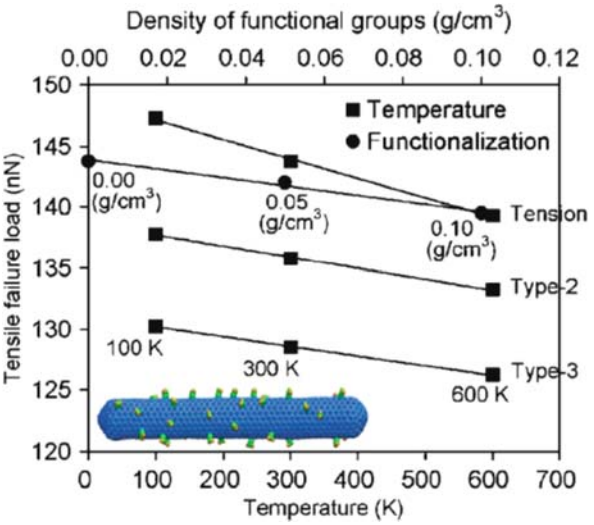


Figure 10-4. Tensile fracture load of hollow SWNTs under tension or combined tension-torsion as a function of either the density of chemical functional groups or temperature. The system temperature is maintained at 300 K in the simulations where the effect of chemical functionalization is investigated. "Type" means the different loading types for combined tension–torsion. The inset illustrates a representative functionalized SWNT where the density of functional groups is 0.10 g/cm^3. After [39] with permission of AIP (2007)

at densities about 1.8–3.6 times higher [49]. Figure 10-4 also indicates that higher temperature decreases the tensile fracture load. However, neither chemical functionalization nor temperature was predicted to influence the tensile elastic moduli of the CNTs.

10.3.1.4. Effect of Combined Loads

In the applications based on CNTs, combined tensile and torsional loads are expected to occur on the CNTs. For instance, CNTs may be used as drive shafts [4], torsion bar springs [7], and torsional actuators [93] that can experience torsion as well as tension. Thus, understanding the mechanical responses of CNTs undergoing this type of loading is important in optimizing their use in new materials and devices. While numerous studies have examined some aspect of the tensile responses of CNTs, such as their strength [38, 41, 44, 90], elastic modulus [43, 84], and twist induced by tension [19] under uniaxial tensile loading, there is much that is still unknown about other aspects of the tensile responses of CNTs in combined tensile and torsional loading.

Jeong et al. [39] examined the tensile responses of hollow, filled, and functionalized single-walled CNTs under combined tensile-torsional loading at various temperatures using MD simulations. They reported that the tensile fracture load decreases linearly with applied torsion and is unaffected by filling the CNTs in so far as torsional buckling does not occur in filled CNT systems. However, if torsional

buckling occurs under combined loading, especially in the case of large degrees of torsion, filling CNTs substantially increased the tensile fracture load. In this case, the tensile fracture load showed notable differences that depended on the type of filling material, because of their differing influence on critical and postbuckling behaviors.

Jeong et al. also investigated the tensile elastic modulus under combined tension-torsion [39]. It was shown that the tensile elastic modulus within the linear elastic limit is almost the same for this system for different loading types. This means that applied torsion does not affect the tensile elastic modulus. These results agree well with the linear elasticity of continuum mechanics theory, where the total strain energy under combined tension-torsion loading is simply the sum of the strain energy for each loading direction, and thus there is no interaction between tension and torsion. This is because the shear strain by applied torsion is much smaller than the tensile strain within the linear elastic limit.

Combined loading also influences the development of failure criteria for CNTs. When stress is applied in a uniaxial manner, stress and strength can be compared directly to estimate whether or not the part will fail. This comparison is relatively simple because there is only one value of stress and strength. However, the problem becomes more complex when the stress state is multiaxial. In such cases there are a multitude of stresses but only one significant value for strength, and this requires that failure be characterized using multiaxial strength (or failure) criteria [82, 83]. Under multiaxial loading conditions, the details of failure at the micromechanical and nanomechanical levels are so incomplete that the failure process cannot be followed analytically. Thus, failure criteria for macroscopic objects have evolved from attempts to develop analytical or empirical macromechanical models to describe experimental observations of failure under multiaxial loading [82, 83]. Such failure criteria use the concept of "a failure surface" or "a failure envelope" generated by plotting principal stress components in principal material axes [82, 83].

Jeong et al. [40] recently developed the failure criteria of CNTs using a multiscale approach that adopts macromechanical or continuum mechanics models (or stress method) [28] to describe computational, atomic-scale observations of fracture in MD simulations. Accordingly, these failure criteria also use the stress method and the concept of "the failure surface" or "the failure envelope" which is generated by plotting principal stress components along the principal material axes in the same manner as is commonly done for macroscopic objects [82, 83]. In atomic-scale systems, the pressure may be calculated by two different methods. One is based on the virial theorem [61] and another is based on the concept of stress from continuum mechanics [82, 83]. Numerous studies have compared these two methods and have shown that they are equivalent for homogeneous systems in equilibrium [94–97]. However, in the case of nonequilibrium or inhomogeneous systems, the stress method remains valid while the virial theorem leads to unphysical results [94–97]. Most previous studies have therefore used the stress method to determine the mechanical stress of CNTs [38, 42]. If the failure criteria of CNTs were derived from a virial stress, the values of the stress might be different from those determined from the mechanical stress, but the trends would be expected to be the same [94–97].

Additionally, if virial stresses were used thermal fluctuations could be determined. However, these fluctuations have been determined to have a negligible effect on systems such as CNTs [38, 42, 94–97]. Consequently, in the work of Jeong et al., the use of the stress method and the exclusion of the virial theorem are acceptable [40].

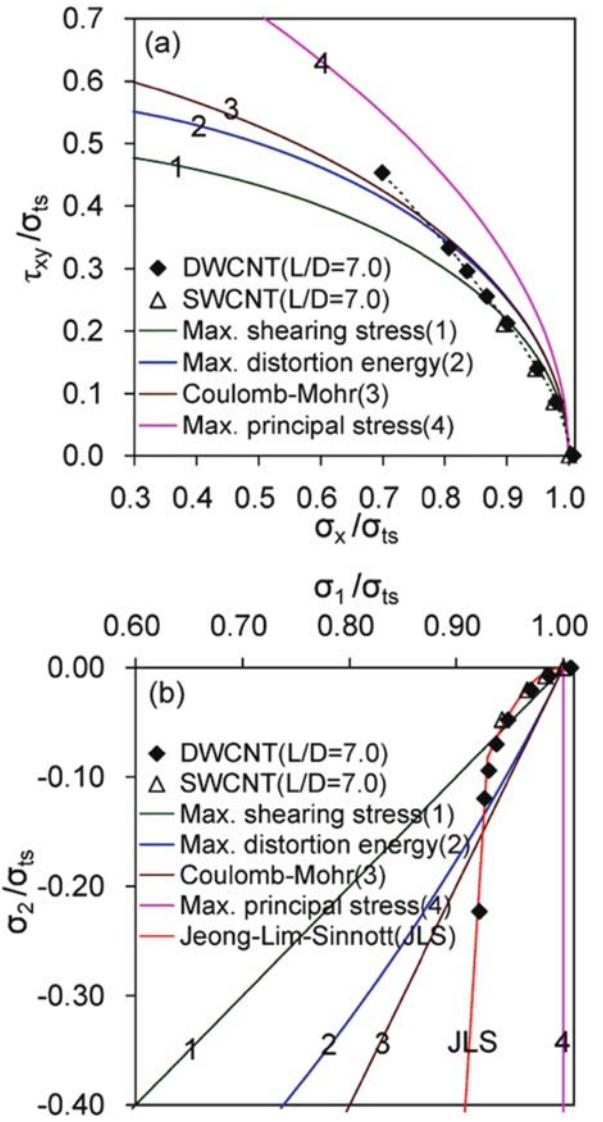


Figure 10-5. Failure envelopes for CNTs and failure envelopes predicted from failure criteria for macroscopic objects. L is the length and D is the diameter. (**a**) Failure envelope generated by τ_{xy} (torsional shear stress) and σ_x (tensile stress), which are normalized with respect to σ_{ts} (tensile strength) under uniaxial tension. (**b**) Failure envelope generated by σ_1 (maximum principal stress) and σ_2 (minimum principal stress), which are normalized with respect to σ_{ts} (tensile strength) under uniaxial tension. After [40] with permission of IOP (2007)

Figure 10-5 [40] indicates that the failure envelopes of CNTs are significantly different from the predictions from the failure criteria for macroscopic objects since these theories do not take into account the different possible failure mechanisms that can occur in nanometer-scale CNTs. The figure also illustrates that the failure strength or envelope of the CNTs under combined loading is importantly different from what occurs under uniaxial tensile loading condition. The interaction failure curves or failure envelopes in Figure 10-5a or b can be used to determine whether CNTs will fail under combined tension–torsion, with the failure criteria expressed by principal stresses [40].

10.3.2. Compressive Behavior

As mentioned earlier, when the CNTs are placed under compression, their behavior is different from their behavior under tension due to the fact that they are much more compliant in the radial direction. This section discusses some of the important insights and findings regarding the response of CNTs to compressive behavior that have been obtained from MD simulations.

10.3.2.1. Buckling Instability

CNTs exhibit structural instability under compression as they are much more compliant in their radial direction than in the axial direction. However, the deformed structure remains within the elastic limit without any plastic yield or bond breaking. Yakobson et al. performed pioneering computational simulations on the compressive buckling behavior of CNTs using MD simulations [1]. Their results indicated that the nanotubes exhibit great flexibility, and may be severely deformed without breaking any chemical bonds. They also developed a continuum shell model to describe the buckling modes of the CNTs. After the pioneering work of Yakobson et al., numerous other computational studies also observed compressive buckling of CNTs. For example, Ozaki et al. [26] predicted from MD simulations that ripple shell buckling occurred and was strongly dependent on temperature. In their study, the stress under large strain and zero temperature varied with helicity. Cornwell et al. [98] also examined the compressive elastic response and critical strain of SWNTs by means of MD simulations. They found that the critical strain varies strongly with tube radius. The critical strain predicted from MD simulations was in agreement with calculations based on continuum elasticity theory for tube radii greater than 10.0 Å. However, for tube radii less than 10.0 Å, the values from two models were diverged.

The compressive buckling of CNTs was also examined by Buehler et al. [17] using MD simulations. In particular, they examined the effect of aspect ratios on the compressive buckling of the nanotubes and reported that the long tubes with large aspect ratios display significantly different mechanical behavior than tubes with smaller aspect ratios. They distinguished three different classes of the response to compressive loading. While the deformation mechanism was characterized by the buckling of thin shells in nanotubes with small aspect ratios, it was replaced by a rod-like buckling mode above a critical aspect ratio, analogous to the Euler

theory in continuum mechanics. For very large aspect ratios, a nanotube was found to behave like a flexible macromolecule which tends to fold due to van der Waals interactions between different parts of the carbon nanotube. This suggests a shell-rod-wire transition of the compressive behavior of carbon nanotubes with increasing aspect ratios.

There have been some instances when different simulation studies predicted different responses for similar CNT systems. For example, Trotter et al. [51] and Ni et al. [36] predicted different average buckling forces and critical strains for compressed nanotubes using the same potential in the simulations. These differences were attributed to differences in the number of rigid moving atoms and the size of thermostat regions. This indicates the importance of the relative thermostat atom configurations and types on the results of MD simulations. Heo et al. examined the influence of thermostat methods, the number of thermostat atoms, and the rate of deformation on the compressive behavior of CNTs [30]. They reported that the compressive buckling force of CNTs can vary with the relative fraction of Langevin thermostat atoms, and nonphysical buckling can occurs during deformation with large relative percentage of thermostat atoms and at higher deformation rates. In contrast, the Nosé-Hoover and velocity rescaling thermostats resulted in physical compressive deformation modes regardless of the percentage of thermostat atoms. In addition, they found that the Langevin and velocity rescaling thermostats are successful at maintaining a constant system temperature during the compression of CNTs, but the Nosé-Hoover is not able to do so and this is therefore not considered to be suitable for use in simulations of the compressive deformation of CNTs. The temperatures of CNTs with 100% thermostat atoms are plotted in Figure 10-6 [30].

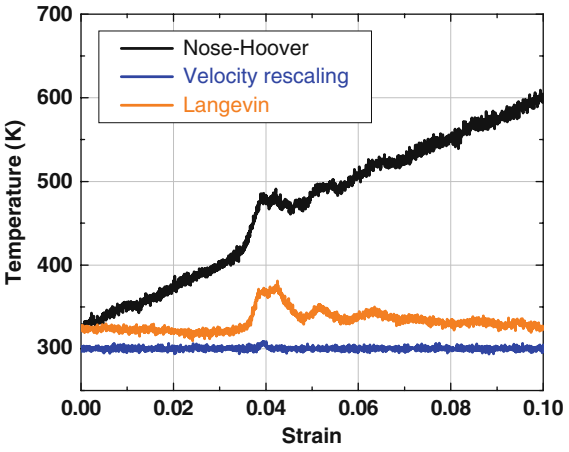


Figure 10-6. System temperature versus strain for the n-butane filled nanotube using the indicated thermostats at a deformation rate of 40 m/s. After [30] with permission of ASP (2007)

10.3.2.2. Effect of Filling, Functionalization, and Temperature

The basic compressive behavior of filled CNTs was investigated by Ni et al. using MD simulations [36]. They predicted that filling CNTs with fullerenes, methane or neon increases the compressive buckling loads and decreases the effect of temperature on the buckling. Similarly, Trotter et al. [51] explored the compressibility of CNTs filled with diamond nanowires, smaller nanotubes, C_{60}, methane, neon, n-C_4H_{10}, or n-C_4H_7 molecules using MD simulations. They predicted that nanowire-filled CNTs and MWNTs exhibit similar compressive responses and that filling CNTs increases their buckling load during compression.

Figure 10-7 illustrates the results obtained by Ni et al. [36]. The figure indicates the effect of the type of filling material on the compressive behavior. Regardless of the filling material, filled CNTs have significantly higher buckling forces compared to empty CNTs. The figure also illustrates that the buckling forces are dependent on the density of the filling materials. In an effort to elucidate the effect of filling density on the buckling force, multiple simulations of the CH4-filled CNT were considered. These simulations showed that the buckling force of the filled CNTs is approximately constant until a critical filling density is reached. When the CNT is filled with a suitable number of atoms or molecules, the buckling force begins to increase with increasing density. They also interpreted the elastic modulus of the CNTs from examination of the curves in Figure 10-7 where the slopes of the curves prior to buckling are approximately equal. Therefore, they reported that filling a CNT does not alter its axial elastic modulus.

In order to increase the strength of adhesion between CNTs and a matrix in composite materials, CNTs are sometimes modified through covalent chemical modification of CNT walls [92]. However, there is concern that since the

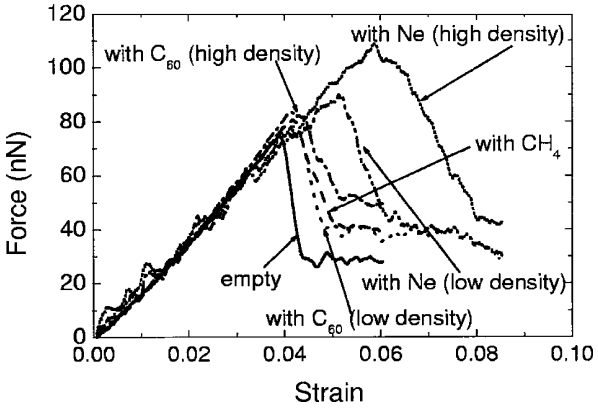


Figure 10-7. Force versus strain for the compression of empty and filled 100 Å CNTs at 300 K (Reprinted with permission from Ni et al. [36]. Copyright (2002) by the American Physical Society. http://link.aps.org/abstract/PRL/v88/p205505)

covalently bonded tethers alter the sp^2-hybridization of CNTs wall at the points of attachment, the compressive properties of the CNTs could be adversely affected. Garg and Sinnott [49] examined the effect of covalent chemical functionalization by $H_2C{=}C$ groups on the compressive buckling behavior using MD simulations. They found that chemical functionalization of nanotube walls leads to slightly lower compressive buckling forces relative to unfunctionalized SWNTs.

Figure 10-8 show the snapshot from simulations of the compression of functionalized CNTs obtained by Garg and Sinnott [49]. The tubule is compressed, kinked, or buckled in the body of the nanotube. Chemical covalent functionalization decreases the buckling force for the tubule by about 19.4%. This means that the functionalized tubule is less stiff in the direction of the tubule axis than the regular unfunctionalized tubule and therefore is predicted to deform more readily. This degradation in stiffness is caused by the formation of sp^3-hybridized carbon "defect" sites on the tubule walls from the attachment of the functional groups. They also reported that helical symmetry has little effect on the results of chemical functionalization but some functional groups dissociate from CNTs with radii below about 0.4 nm during compression due to the high strain associated with these systems.

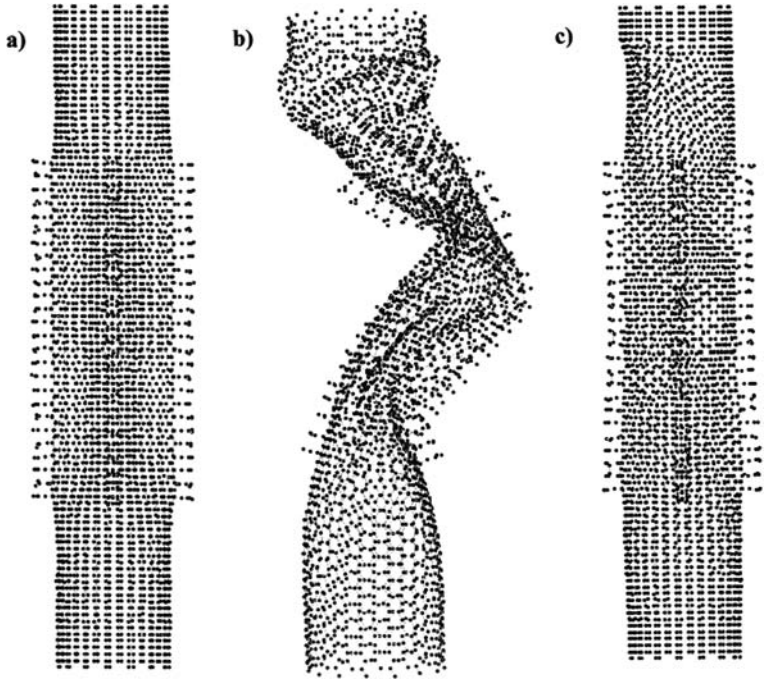


Figure 10-8. Snapshots from the simulation where functionalized carbon nanotubes are compressed. After [49] with permission of Elsevier (1998)

10.3.2.3. Nanotube Proximal Probe Tips

CNTs are sometimes used as proximal probe tips because of their high axial modulus, controlled size, and ability to conduct electrically [2]. Such tips can be used to image fine details on surfaces [8, 14] and, when a current is running between the tip and sample, to etch the sample [99]. Grag et al. and Harrison et al. modeled the compressive behavior that occurs when the CNT comes into contact with a surface using MD simulations [47, 48, 100]. When CNTs indented a surface, the first response was the compressive deformation of the tube cap and then they developed the compressive buckling along the tube length [47]. This behavior is similar to the compressive rod-buckling described by Euler theory of continuum mechanics due to high aspect ratios of the nanotubes [1, 17, 101]. Finally, the end of the CNT slips on the surface [47], as shown in Figure 10-9, if the surface is passivated. The MD simulations showed that these deformations occur elastically regardless of the rigidity of the surface being indented. In addition, similar mechanisms occurred during indentation with a small MWNT and a bundle of SWNTs [48]. In the case of MWNT, it was found that while shell-shell interactions have little effect on the deformation mechanisms, the multiwalled tubule is significantly stiffer than comparably sized SWNTs [48]. Finally, Dzegilenko et al. showed how CNT proximal probe tips can easily etch or penetrate the Si(001) surface during indentation even without the presence of a current between the tip and the surface using MD simulations [3]. To summarize, MD simulations reveal the mechanical properties of CNTs and surfaces that are most important for nanometer-scale indentation. The insights gained from these simulations help in the interpretation of experimental data.

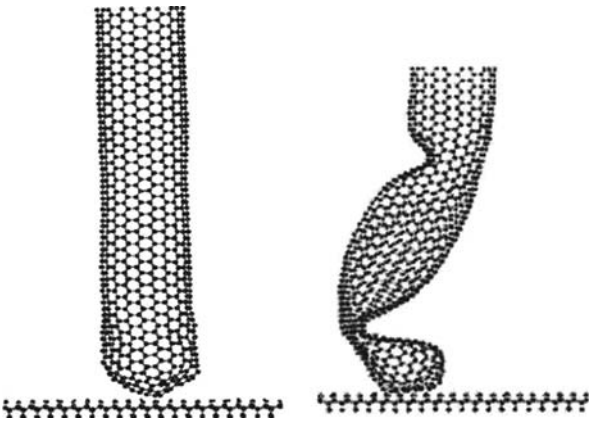


Figure 10-9. Snapshots from a molecular dynamics simulation of a (10,10) SWNT indenting hydrogen-terminated diamond (111). (Reprinted with permission from Garg et al. [47]. Copyright (1998) by the American Physical Society. http://prola.aps.org/abstract/PRL/v81/i11/p2260)

10.3.2.4. Crystalline Bundle

Typically, CNTs are self-organized into crystalline bundles to maximize their van der Waals interactions. As CNT bundles are difficult to disperse, almost all of the fibers in experimental nanotube composite materials [15] are likely to be bundles, and therefore their mechanical responses should be investigated to optimize their use in these applications. MD simulations have been commonly used to investigate the compressive behavior of CNT bundles and the influence of the bundle environment as in the works of Liew et al. [24], Jeong et al. [32], and Grag et al. [48]. Liew et al. [24] found that the long-range van der Waals interactions improve the overall rigidity, buckling load, and critical strain of the CNT bundle depending on the bundle size. Jeong et al. [32] also examined the effect of the intertube van der Waals interactions within the bundle and demonstrated that the intertube van der Waals interactions apparently increase the buckling load and stiffness of the bundle. Grag et al. [48] examined the compressive behavior of a small CNT rope indented against diamond and graphene to assess the effect of intertubule interactions on deformation. They simulations revealed how the deformation of the rope leads to the distortion of its end and allow for the determination of the effect of shear stress within the bundle on the buckling force of the rope. They reported that intrabundle interactions due to shear forces play a critical role in determining the maximum buckling force of the CNT rope tip. To summarize, the MD simulations revealed that the intertube van der Waals interactions within the CNT bundle apparently make an important contribution in improving the compressive properties of the bundle.

10.3.3. Bending Behavior

If CNTs are used as key elements of nanoelectromechanical systems (NEMS) such as bendable nano-oscillators and actuators [10], then their bending and oscillatory properties will be especially important. This section discusses some of the important insights and findings on the bending behavior of the CNTs that have been obtained from MD simulations.

10.3.3.1. Bending Modulus

The measurement of the bending modulus of CNTs is made by assuming the CNTs to be elastic beams, and thus requires determination of CNT volume and effective wall thickness, as was the case for tensile deformations. Srivastava et al. [43] predicted that the bending modulus of small diameter SWNTs is about 0.9 TPa using MD simulations in agreement with the value measured experimentally [10]. In addition, they found that the bending modulus decreases as the diameter of the CNTs increases. Poncharal et al. [10] observed that, while small tubes exhibit a modulus of around 1 TPa, larger tubes are much more compliant with a modulus of around 0.1 TPa. This dramatic reduction of the stiffness is attributed to the so-called rippling effect which reveals nearly periodic wavelike distortions in the compressed section of the tube [16].

10.3.3.2. Buckling Instability

CNTs can be deformed easily under bending load because of their compliance in the radial direction. When the CNTs are returned to their original positions, there is no plastic deformation as a result of the bending [1, 22]. This is a remarkably elastic phenomenon that would not take place in other graphitic structures and is indicative of the structural instability of the CNTs under bending loads. Numerous studies have been performed to investigate the bending buckling behavior of CNTs using MD simulations. For example, Ijima et al. [22] found that atomistic simulations of the responses of nanotubes to bending are comparable to the images observed in high resolution electron microscope. Yakobson et al. [1] also examined the bending responses of SWNTs. Their results indicated that nanotubes may be severely bent without breaking any chemical bonds. Cao and Chen [18] estimated the bending buckling strain and curvature of larger nanotubes (with radii up to 27 Å) using both MD and finite element modeling. These studies identified a single critical discontinuity between bending and buckling of nanotubes, which is delineated by a change in the functional dependence of the strain energy of the nanotube as a function of the bending angle [23].

10.3.3.3. Effect of Filling, Functionalization, and Temperature

Filling CNTs can affect their bending response. Danailov et al. [20] used MD simulations to predict that filling CNTs with Au nanowires would increase the bending buckling force and that nanowire-filled CNTs and MWNTs exhibit similar bending responses. They also predicted that filling CNTs increases their stiffness during bending. This is different from the cases of tensile, compressive, and torsional loading, where filling CNTs did not alter their stiffness [33, 36, 39]. The influence of filling on CNT bending was also investigated by Heo et al. [102]. They examined the bending responses of pristine hollow, C_{60}-filled, n-butane-filled, and multi-(dual-, triple- and quadruple-) walled CNTs at various temperatures using MD simulations, and then compared the results with the responses of CNTs with covalently bonded functional groups and wall vacancies. The results indicated that filling CNTs increases their bending buckling forces and that MWNTs support higher bending buckling loads than SWNTs and peapods because of the presence of the inner CNTs. Their simulations also reported that the bending behavior of hollow or filled CNTs can deteriorate at high temperature.

During the synthesis or purification of CNTs, defects such as vacancies can be introduced into the nanotube walls. CNTs also can be treated to covalently attach chemical groups to the sides [103] or ends [13] and the site of the functional group attachment can act like wall defects. This is because the sp^2-hybridization of the carbon atoms around the defects and functional groups are altered. Heo et al. [102] investigated the effect of wall defects (vacancies or $H_2C{=}C$ functional groups) on the bending responses of the CNTs. Figure 10-10 [102] illustrates that the maximum bending forces decreases with the number of wall defects and the density of the functional groups, which is very similar to the defect effect on the compressive properties of nanotubes [49, 50]. In summary, they found that the chemical

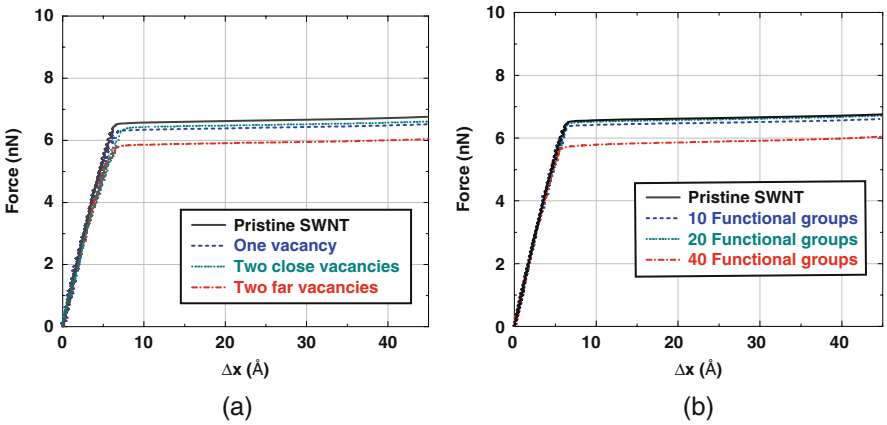


Figure 10-10. Force versus deflection of 100 Å (10,10) hollow pristine CNT and CNTs having various densities of (**a**) vacancy defects and (**b**) functional groups during bending

modification of the CNT wall through the generation of vacancies or the attachment of functional groups decreases the bending buckling forces of CNTs [102].

10.3.3.4. Effect of External Gases

Nanotubes are proposed for use in applications such as nanobalance [10] and nanovalve [6] devices. It was predicted that the nanobalance can weigh particles in the femtogram to picogram range [10] and the nanovalve can control the flow rate of fluid through nanometer-scale channels [6]. There were also attempts to use the nanotubes as electromechanical oscillators [11]. "Guitar-string-like oscillation modes" of CNTs suspended between two contacts were detected and tuned. These CNT-based NEMS devices can be exposed to an external gas fluid that affects the bending responses of the nanotubes. Therefore, understanding these bending responses in response to external gas interactions is important to optimize their use in the NEMS devices and to determine how the external gas influences the bending compared to the responses discussed above in perfect vacuum.

Heo et al. [104] investigated the fundamental vibrational properties of bridged and cantilevered CNTs responding to external gases, such as Ar, Kr, and Xe, using MD simulations. In the vibrational simulation of the bridged CNT, both ends of CNT were fixed and the other part of the CNT as well as the noble gas atoms moved free following Newton's equation of motion, as shown in Figure 10-11a. The center region of the CNT was then pulled out by 0.3 nm which is the initial bending displacement, by applying additional force to the atoms in that region. For the cantilevered CNT simulations, atoms in one end of CNT were set to be rigid atoms and then the other part of the CNT and the fluid molecules were changed into active atoms as indicated in Figure 10-11b. During the simulations, the deflection of CNT was measured, which allows for the calculation of the vibrational frequency. The

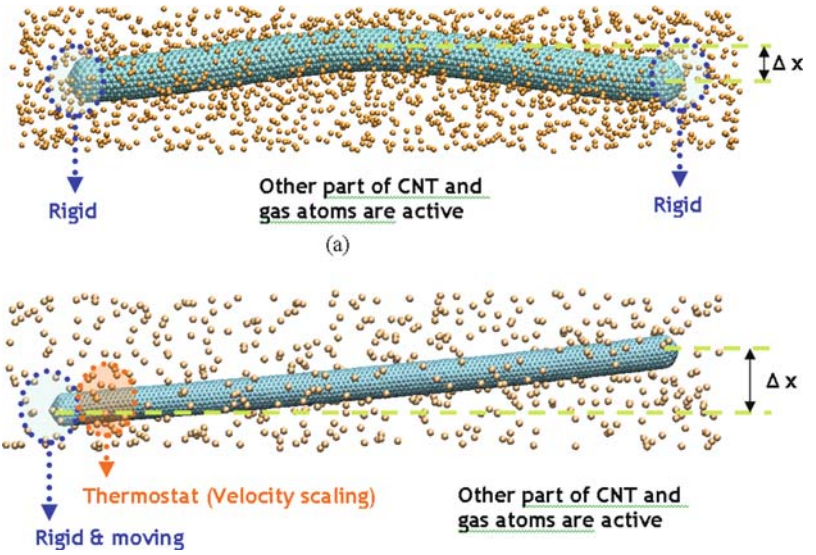


Figure 10-11. Snapshots from the MD simulations of (**a**) a bridged 20 nm long (10,10) hollow single-walled CNT in the 100 atm Ar fluid and (**b**) a cantilevered 30 nm long (10,10) hollow single-walled CNT in the 10 atm Kr fluid. Carbon atoms are in shown in *blue*, and the Ar and Kr atoms are shown in orange

results indicated that the CNT vibrational frequencies depend on the CNT length as well as the pressure and mass of the fluid atoms.

Lee et al. [52] also examined the bending responses of CNTs exposed to an external gas fluid. Specially, they investigated the responses of single and multiwalled nanotubes to impacts with noble gas atoms using MD simulations to predict the motion of nanotubes when they are used for the devices located in the path of pulsed fluid flow. In their simulations, it was predicted that the more flexible SWNTs buckle over during the relaxation stage that follows the collision events. They also found that as the number of collisions and the number of walls increase, the amplitude of nanotube oscillations increases, and that as the number of walls increases, the oscillations of the MWNTs are balanced in the upward and downward directions. The deflections of zigzag and armchair nanotubes were compared for similar numbers of walls and nanotube diameters. As the nanotubes are shortened, the vibrational motion of the nanotubes was predicted to be damped by energy dissipation. This work was an important first step to understanding their response to external fluid flow, which is likely to influence the behavior of nanotube levers in applications such as NEMS. In addition, understanding the oscillatory deflection of nanotubes is also important in applications such as nanoactuators, nanoswitches, and nanotweezers, where large displacements are repeatedly induced. Figure 10-12a contains a typical snapshot after the first Ar collision event onto the SWNT. After 10 collision events, the nanotube bends and "rumples" form in the wall structure, as indicated in Figure 10-12b. The SWNT, which is more flexible than the double-walled carbon

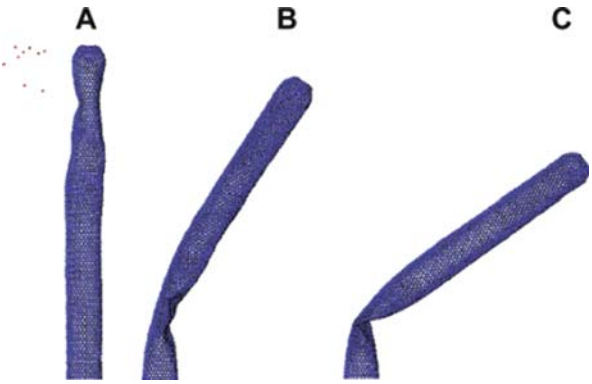


Figure 10-12. Snapshots of a (28,0) SWNT after a series of Ar atom collisions and during subsequent relaxation. (**a**) The nanotube after the first collision event. (**b**) The nanotube after the tenth collision event. (**c**) The nanotube after relaxing for 40 ps. The left-most end was held rigid throughout as described in the text. After [52] with permission of ACS (2005)

nanotubes (DWNTs) and triple-walled carbon nanotubes (TWNTs), even buckles over during relaxation, as shown in Figure 10-12c.

10.3.4. Torsional Behavior

As stated above, the unique properties of CNTs make them attractive for use in applications such as NEMS [4, 7, 9, 12, 15]. In these and similar applications, it is expected that torsional loading will occur on the CNTs. For instance, the CNTs may act as torsional springs in resonators [9] or act as rotational bearings in actuators [4]. In these applications, the torsional and oscillatory properties of the CNTs significantly affect the performance of the devices. Thus, understanding the torsional responses of CNTs is important to optimize their use in new devices.

10.3.4.1. Shear Modulus and Stiffness

As mentioned earlier, the measurement of the shear modulus of CNTs is also made by assuming the CNTs to be elastic beams and thus requires determination of CNT volume. Equation (10-13) is the definition of shear modulus based on continuum mechanics theory. Therefore, the shear modulus of CNTs can be calculated from their strain energy and volume in an analogous manner.

$$G = \frac{2U}{V\gamma^2} = \frac{2U}{AL\gamma^2} \tag{10-13}$$

where V is the volume of the body, A is the cross-sectional area, L is length, U is the strain energy, and γ is the shear strain. The cross sectional area of the CNTs can be calculated from the effective wall thickness of 0.34 nm [84].

The shear modulus of various nanotubes has been predicted, most commonly from atomistic calculation methods [33, 43, 84]. Several experimental approaches

have also been developed and used to measure the shear modulus of the nanotubes [105, 106]. As described earlier, MD simulation method is simulated at high strain-rates due to the limitation in the time scale of the phenomenon. This high strain-rate significantly affects the torsional buckling loads of CNTs but not the torsional stiffness or shear modulus [31]. Therefore, the torsional stiffness or shear modulus can be calculated directly using the MD simulation method, but care must be taken in interpreting the results and comparing them to one another and to experimental data.

Some studies have been performed to determine the shear modulus of CNTs. In most studies, the shear modulus of CNTs has been predicted to be about 0.35–0.45 TPa [33, 43, 84, 106]. Lu [84] predicted a value of about 0.45 TPa using an empirical force constant model. Similarly, Srivastva et al. reported a value of about 0.35 TPa using classical MD simulations [43]. Jeong et al. found that the torsional shear modulus of the CNTs is about 0.34 TPa using MD simulations [33]. In their work, it was founded that the torsional shear modulus is relatively independent of CNT diameter and length in the same way as in continuum mechanics theory. Recently, Hall et al. [106] experimentally measured a value of 0.4 TPa using a CNT-based torsion spring.

The torsional stiffness, K, of a body twisted by external torsion can be expressed as

$$K = L\frac{d^2U}{d\theta^2} \tag{10-14}$$

where L is the length, U is the strain energy, and θ is the torsional angle. Recently, Jeong et al. [33] calculated the torsional stiffness of CNTs using Eq. (10-14) and MD simulations, to determine the dependence of the torsional stiffness on CNT diameter and length. Figure 10-13 presents the torsional stiffness of hollow CNTs in terms

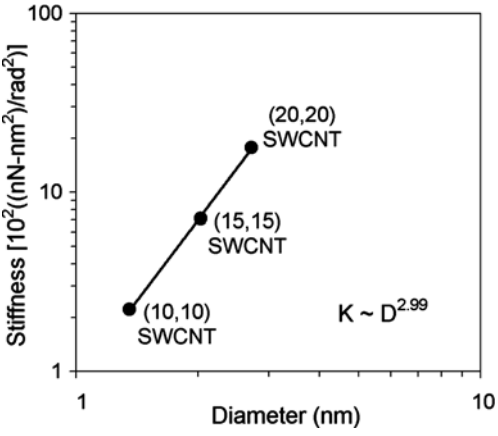


Figure 10-13. Torsional stiffness of the CNTs in terms of their diameters. K and D are the torsional stiffness and the diameter of CNTs, respectively. After [33] with permission of AIP (2007)

of their diameters [33]. The dependence of the torsional stiffness on the diameter of CNTs is predicted to vary as $K \sim D^{2.99}$, where D is the nanotube diameter. This is in good agreement with the prediction of the linear elasticity of continuum mechanics theory, where the strain energy of a thin-walled cylinder can be expressed as:

$$U = \frac{1}{2} G \iiint \gamma_{\text{torsion}}^2 dV = G\pi h R^3 \frac{\theta^2}{L} \tag{10-15}$$

where G is the torsional shear modulus, γ_{torsion} is the torsional shear strain, h is the thickness, R is the radius, θ is the torsional angle, and L is the length. From Eqs (10-14) and (10-15), the torsional stiffness of a thin-walled hollow cylinder or tube can be expressed as.

$$K = L \frac{dU^2}{d\theta^2} = 2G\pi h R^3 \tag{10-16}$$

From Eq. (10-16) we can see that the torsional stiffness of the thin-walled cylinder varies as $K \sim R^3$ (or $K \sim D^3$), in excellent agreement with the results predicted from these MD simulations.

Jeong et al. also examined the dependence of the torsional stiffness on the CNT length using MD simulations [33]. They found that the dependence of the torsional stiffness per unit length on the CNT length varies as $K' \sim L^{-0.99}$, where K' is the torsional stiffness per unit length and L is the CNT length. This dependence is in agreement with the definition of the torsional stiffness ($K' = T/\theta = K/L$, where K is the torsional stiffness).

10.3.4.2. Buckling Instability

As mentioned above, when the CNTs are placed under torsion, buckling instability can occur in the same manner as their responses under compressive or bending loads [1]. The CNTs under torsion show that their stiffness is greatly reduced after a certain torsional angle is achieved. However, despite the rapid reduction of torsional stiffness of the nanotubes, no plastic yield or bond breaking occurs for them. This indicates that the CNTs remain highly elastic and the predicted torsional behavior is due to torsional-buckling instability. The bifurcation at which the torsional stiffness of CNTs starts to sharply decrease is the torsional buckling point.

Yakobson et al. performed pioneering computational work on the torsional buckling behavior of CNTs [1]. In particular, they modeled torsional buckling of CNTs using MD simulations and found that the increase of azimuthal angle φ between the tube ends results in abrupt changes of energy and morphology. In their study, the nanotubes were predicted to be remarkably resilient up to extremely high strain with no plastic deformation or atomic rearrangement. They also developed a continuum shell model to describe the torsional buckling modes of the CNTs. There have been other studies on the torsional responses of CNTs using computational

[33, 43, 84] and experimental [105, 106] methods. However, there is much less data available on the torsional buckling of CNTs compared to the data available on the compressive or bending buckling of the nanotubes.

Since the pioneering work of Yakobson et al., MD simulations have been commonly used to investigate the torsional behavior of CNTs. For example, Jeong et al. [33] examined the torsional responses of several different nanotubes at various temperatures using MD simulations, and the results were interpreted and compared to the predictions of continuum mechanics theory. They considered how the torsional responses of various CNTs, such as the critical torsional moment and stiffness, are influenced by the CNT conditions. They reported that the torsional buckling direction of the CNT walls occurs at about 45° of the longitudinal axis of the CNTs. According to continuum mechanics theory, the inclined torsional buckling is caused by shear loading of the walls of twisted tubes, which gives rise to compressive buckling stresses at 45° to the longitudinal direction as predicted by Mohr's circles [101]. Therefore, it is interesting that the torsional buckling deformation of nanoscale CNTs agrees so well with continuum mechanics theory. They also compared the torsional buckling moment between SWNTs and MWNTs in terms of their diameters [33]. The results indicated that the torsional buckling moment in MWNTs increases about 3.3 more rapidly than the SWNTs when their diameters increase and this is mainly caused by the presence and number of inner tubes in MWNTs.

10.3.4.3. Effect of Filling, Functionalization, and Temperature

There has also been considerable interest in filling CNTs with various materials [91] and in chemically modifying CNTs [92, 107], which can affect their mechanical response. Jeong et al. [33] investigated the effect of functionalization and filling on the torsional behavior at various temperatures using MD simulations. In their study, CNTs were filled with n-butane or C_{60} of which the densities are 0.53 and 1.21 g/cm^3, respectively. In addition, CNTs that have $H_2C{=}C$ groups covalently and randomly bonded to the CNT walls were also considered to address the influence of chemical functionalization. The density of functional groups considered was 0.05 g/cm^3 (or a CNT/$H_2C{=}C$ weight ratio of 35.08) and 0.10 g/cm^3 (or a CNT/$H_2C{=}C$ weight ratio of 17.54), a value that is slightly lower than those reported experimentally [92]. Their simulations indicated that filling and functionalization of CNTs importantly change the torsional buckling moment but not the shear modulus.

Figures 10-14 and 10-15 summarize the findings of Jeong et al. on the effect of CNT filling [33]. Figure 10-14 indicates that there are significant differences in the torsional buckling moments between hollow and filled CNTs. This is due to the torsional buckling characteristics of the CNTs that are affected by the presence of the filling material. The torsional buckling of macroscopic thin-walled tubes is indicated by the growth of bulges, waves or ripples that contract and expand the tubes in the radial direction. As can be seen in Figure 10-15, the torsional buckling of CNTs also contracts and expands the tubes in the radial direction in the same way as the torsional buckling behavior of macroscopic thin-walled tubes. However, in the case of

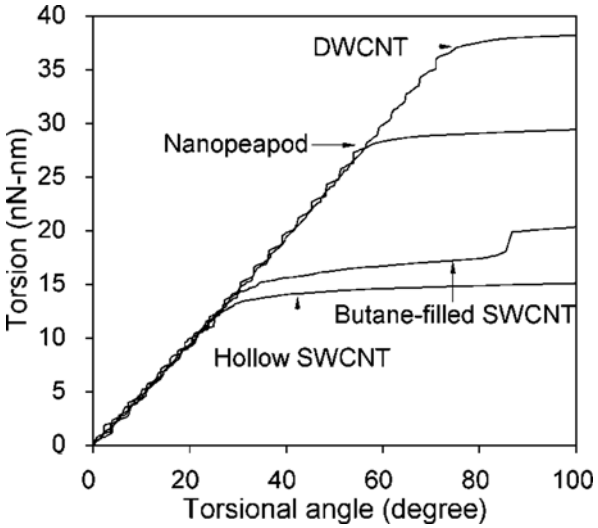


Figure 10-14. Torsion-torsional angle curves of hollow and filled CNTs. After [33] with permission of AIP (2007)

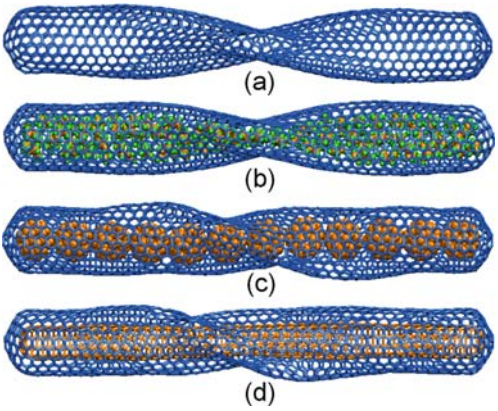


Figure 10-15. Buckled configurations of hollow and filled CNTs; (**a**) Hollow CNT. (**b**) Butane-filled CNT. (**c**) Nanopeapod. (**d**) (05,05) SWNT-filled CNT. After [33] with permission of AIP (2007)

the filled CNTs, the radial-direction contraction decreases, leading to increased van der Waals interaction forces between the CNTs and filling materials that influence the progress of torsional shear strain. Consequently, the torsional buckling moment can be increased by the presence of filling materials, and that the amount of this increase depends on the kind of filling materials. In addition, filling CNTs under torsion may raise the rate of increase in the force needed to buckle CNT systems more than under bending and compression loading [36, 102]. Figure 10-14 further indicates that the torsional stiffness and shear modulus are unaffected by the presence of filling materials due to weak van der Waals shear interactions. This finding

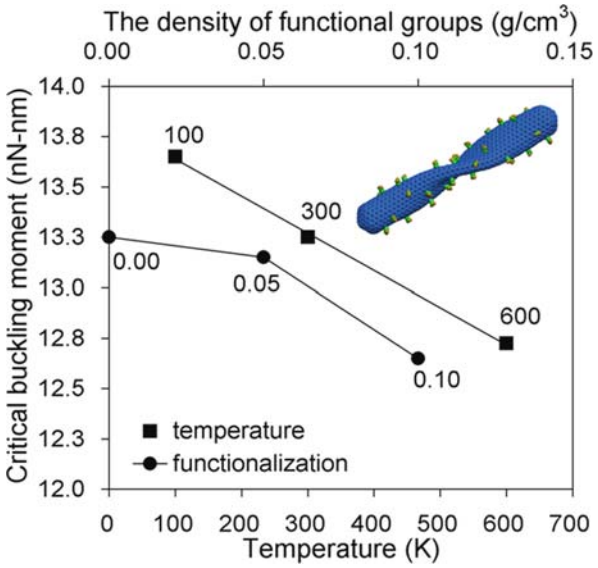


Figure 10-16. Torsional buckling moment of hollow (10,10) CNTs in terms of the density of chemical functional groups and temperature. The inset illustrates Buckled configuration of chemical functionalized (10,10) CNTs. After [33] with permission of AIP (2007)

is in agreement with the results of studies on the compression and tension of filled CNTs [36, 39, 51]. Thus, filling CNTs does not alter their torsional shear modulus as well as axial elastic modulus. However, the bending elastic modulus is predicted to be affected by filling CNTs [20].

In this work of Jeong et al. [33], the effect of covalent chemical functionalization by $H_2C{=}C$ groups on the torsional buckling behavior of (10,10) CNTs was investigated by MD simulations. In Figure 10-16 [33], we can see that this type of chemical functionalization decreases the torsional buckling moment and its decrement depends on the density of the chemical functional groups. Specifically, the rate of decrement is about 0.8 and 4.5%, respectively. These values differ little from the 1.2 and 3.0% decrease predicted for tension of (10,10) CNTs functionalized with the same groups and densities [39], but are lower than the 12.5% decrease predicted for compression of (10,10) CNTs functionalized with the same groups, but at densities that were about 1.8–3.6 times higher [49].

As it is known that temperature influences the mechanical behavior of CNTs, the effect of temperature on the torsional buckling behavior of (10,10) CNTs was also examined by Jeong et al. [33]. Figure 10-16 illustrates that higher temperatures decrease the torsional buckling moment [33]. Particularly, the rate of decrement is about 2.9 and 6.8%, respectively in comparison with values at 100 K. These values differ little from the 2.4 and 5.4% decrease predicted for tension of (10,10) CNTs maintained with the same temperature [39]. However, neither chemical functionalization nor temperature was predicted to affect the CNT torsional stiffness and shear modulus [33].

10.3.4.4. Effect of Combined Loads

In the NEMS based on CNTs such as actuators [4], springs [12], and oscillators [9], a combination of tensile and torsional loads can occur on the CNTs and produce mechanical responses that differ substantially from uniaxial loading [39, 101]. Thus, understanding all aspects of the mechanical responses of CNTs under this type of combined loading is important to optimize their use in new devices.

In an examination of the tensile responses of CNTs under a combination of tension and torsion, it was showed that the tensile elastic (or Young's) modulus is unaffected by the applied torsion, in agreement with linear elasticity theory [39]. This is because the torsional shear strain energy is much smaller than the tensile strain energy and thus does not influence the progress of tensile strain [39]. In addition, the tensile fracture load was decreased by the application of torsion [39]. It is possible, however, that the effects of tension on the torsional responses of CNTs may be significantly different from the effects of torsion on the tensile responses mentioned above. For instance, the torsional buckling moment may be increased by the presence of tension, which can delay the onset of torsional buckling [101], and the torsional shear modulus may be influenced by increases in tensile strain energy. Here, Jeong et al. [34] examined the effects of tension on torsional responses, including the torsional buckling moment and shear modulus, of hollow and filled CNTs under combined tensile-torsional loading using MD simulations.

Figure 10-17 shows the effects of applied tension on torsional properties in terms of an applied tension ratio F_x/F_{ts} [34]. Here, F_x is the tensile load applied simultaneously with torsion and F_{ts} is the fracture load under uniaxial tension. Figure 10-17a indicates how much the critical torsional moment T_{cr} increases under a combination of tensile and torsional loading. The figure indicates that the increment rate $T_{cr}/T_{cr\text{-}pure}$, where $T_{cr\text{-}pure}$ is the critical torsional moment under pure

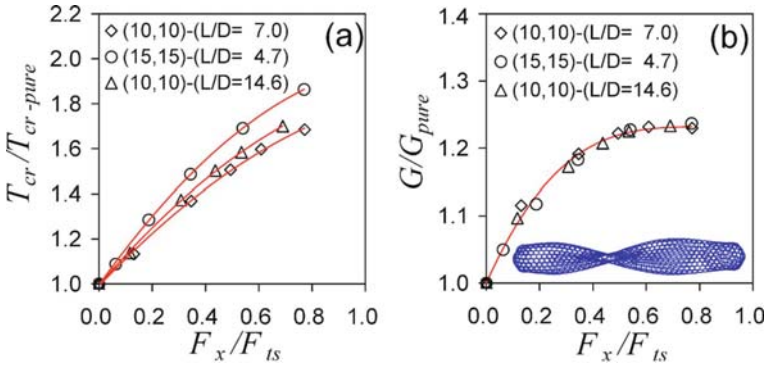


Figure 10-17. Variation curves of the critical torsional moment T_{cr} and shear modulus G of hollow SWNTs under combined tension-torsion in terms of an applied tension ratio F_x/F_{ts}. The ratio *L/D* refers to the nanotube aspect ratio. (**a**) Increment rate $T_{cr}/T_{cr\text{-}pure}$ in the critical torsional moment. (**b**) Increment rate G/G_{pure} in the torsional shear modulus. The inset illustrates the buckled configuration of a (10,10) SWNT with an aspect ratio of 7.0. After [34] with permission of AIP (2007)

torsion, significantly increases in proportion to F_x/F_{ts}. This result is similar to the behavior of macroscopic tubes [101] and is the opposite of the trend for tensile fracture load, which decreases when tension is combined with torsion [39]. Figure 10-17a also shows that $T_{cr}/T_{cr\text{-}pure}$ increases in proportion to the CNT diameter and length.

According to continuum mechanics theory, the torsional shear modulus $G \propto K$, where K is the torsional stiffness [83]. If this is applied to nanotubes, their increment rate in torsional shear modulus and stiffness under combined tension-torsion has the relationship $G/G_{pure} \propto K/K_{pure}$, where G_{pure} *and* K_{pure} are torsional shear modulus and stiffness, respectively, under pure torsional loading. Figure 10-17b illustrates that the torsional shear modulus significantly increases under combined tension-torsion relative to what happens under pure torsional loading [34]. The figure indicates that G/G_{pure} increases in proportion to F_x/F_{ts}, but is unaffected by the geometry of the CNTs. The simulations thus predict that the maximum value of torsional shear modulus under combined loading is 23% larger than for pure torsion. These results suggest that the increase of torsional stiffness by combining tensile and torsional loading may be used to optimize torsional NEMS devices because torsional stiffness is directly related to torsional oscillation frequency; this is a key performance factor of many NEMS devices. For instance, the resonance frequency for torsional oscillation may be expressed as $f = \left[1/(2\pi)\right]\left(K'/I\right)^{1/2}$, where I is the moment of inertia. Therefore, the coupling effect on torsional responses under combined loading should also increase the torsional oscillation frequency of CNTs.

Jeong et al. also investigated the effect of filling CNTs on torsional responses under a combination of tensile and torsional loading [34]. Their study showed that $T_{cr}/T_{cr\text{-}pure}$ for filled CNTs significantly increases in proportion to F_x/F_{ts}, as was the case for the hollow CNTs in Figure 10-17a. They also reported that the torsional shear moduli of filled CNTs under the combination of tension and torsion are higher than those under pure torsion, as was shown to be the case for the hollow CNTs in Figure 10-17b. Interestingly, the study indicated that the values of G/G_{pure} for filled CNTs are not significantly different from the values for hollow CNTs, and they do not depend on the type of filling materials. This is most likely due to the relatively weak van der Waals shear interactions between the CNT walls and the filling materials.

Continuum mechanics, in general, and linear elasticity, in particular, has been applied to study the mechanical behavior of CNTs, either on their own or in combination with molecular mechanics or atomistic simulations [10, 27, 108]. In linear elasticity, the total strain energy under combined tensile and torsional loading is expressed as the sum of the strain energy for each loading direction, because torsion does not work in the tensile direction and tension does not work in the direction of torsion. This means that there is no interaction between tension and torsion, and so the fact that combined loading is taking place does not influence the torsional shear modulus or the tensile modulus. The study of Jeong et al. on the tensile behavior indicated that the tensile modulus of CNTs under combined tension-torsion is unaffected by the applied torsion in agreement with linear elasticity theory [39]. However, as mentioned above, the torsional shear modulus of CNTs under this type

of combined loading is significantly changed by applied tension [34]. This indicates that mechanical coupling influences torsional responses, which is probably caused primarily by tensile strain energy that is much larger than torsional shear strain energy [109]. Accordingly, the torsional shear modulus of CNTs under combined loading can be expressed as:

$$G = G_{\text{pure}} + \Delta G_{\text{coupled}} \tag{10-17}$$

where $\Delta G_{\text{coupled}}$ is the torsional shear modulus increased by the coupling with tension. Linear elasticity interprets the mechanics of materials in terms of unchangeable elastic moduli, and thus the torsional responses of CNTs under combined loading are beyond linear elasticity theory [83].

In some objects there are preexisting stresses before an external load is applied and this initial stress is termed the prestress. Prestressed objects include familiar structures such as rope hammocks and spider webs, and are also exemplified in engineering applications such as prestressed concrete [110], piezoelectric actuators [111], and flexure elements of microelectromechanical systems (MEMS) [112]. In particular, in applications that involve torsional flexure, an axial prestress affects the resulting torsional properties such as torsional strength (or buckling moment) and stiffness [112]. Here, the effect of torsional stiffness is of primary importance because this quantity is directly related to the oscillation frequency, which is the key performance factor of such systems.

Some studies have shown that although an axial prestress affects the torsional stiffness of linear elastic materials, the effect is most important for sections with low torsional rigidity, such as thin rectangular and opened thin-walled sections, and is less important for closed cross sections having relatively high torsional rigidity [112, 113]. However, it has been shown that hyperelastic or large deformation (or nonlinear) elastic materials, such as natural rubbers [114] and biomaterials [115], behave differently from linear elastic materials due to their substantial large axial elongation properties. As a result, their torsional stiffness is altered by axial prestress even in the case of closed circular sections [114, 115].

While CNTs resemble closed, circular, thin-walled objects and have high torsional rigidity relative to their diameter [33, 84, 105, 106, 116], they also have the ability to elongate in the axial direction to a substantial degree [41] relative to linear elastic materials. Hence, it is not clear a priori whether the torsional stiffness of CNTs will be affected by the presence of an axial prestress despite the fact that numerous studies have examined some aspect of the torsional responses of CNTs under pure torsion using theoretical [33, 43, 84] or experimental [105, 106] methods. Here, Jeong et al. investigated the effects of tensile or compressive prestress on the torsional responses, including torsional stiffness and torsional buckling moment using MD simulations [35].

Figure 10-18 illustrates the effects of the torsional stiffness K_u by tensile prestress σ_{tp} (or preload F_{tp}) in terms of the tensile prestress ratio R_{tp} [35]. The tensile prestress was kept below the tensile strength of the CNTs, which is calculated to be about 100 GPa in agreement with previously published theoretical predictions

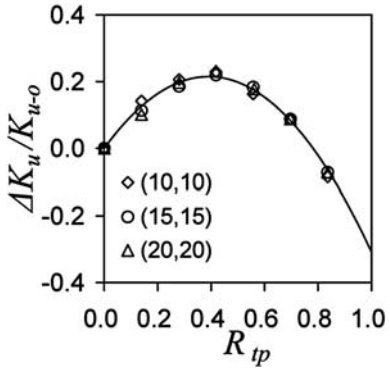


Figure 10-18. Variation rate $\Delta K_u/K_{u-o}$ of the torsional stiffness K_u by tensile prestress σ_{tp} in terms of tensile prestress ratio R_{tp}, which is the ratio σ_{tp}/σ_{ts} of tensile prestress and strength or the ratio F_{tp}/F_{ts} of tensile preload and fracture load. The variation ΔK_u is calculated from the definition of $\Delta K_u = K_u - K_{u-o}$, where K_u and K_{u-o} are the torsional stiffness in the presence and in the absence of tensile prestress, respectively. After [35] with permission of AIP (2008)

[39, 42]. The figure indicates that when ΔK_u is normalized with respect to the torsional stiffness K_{u-o} in the absence of tensile prestress, the variation rate $\Delta K_u/K_{u-o}$ depends on only the R_{tp} and not on CNT curvature, k. In Figure 10-18, it should be noted that the torsional stiffness decreases as the tensile prestress increases above R_{tp}=0.42, and, in particular, is smaller than the torsional stiffness in the absence of tensile prestress above R_{tp}=0.77. Previous studies have shown that the linear elastic limit of (10,10) CNTs under pure tensile loading is about 60 nN [39]. This value is not significantly different from the tensile preload of the (10,10) CNT at R_{tp}=0.42, at which the slope of $\Delta K_u/K_{u-o}$ changes. Accordingly, this unusual result is thought to be caused mainly by the ability of the CNTs to endure large axial elongation before they are broken beyond their linear elastic limit [39, 42]. The torsional stiffness was increased to approximately 23% at R_{tp}=0.42, which is where the maximum $\Delta K_u/K_{u-o}$ occurs. The effect of the tensile prestress was further extended to consider its influence on the critical torsional moment T_{cr} [35]. The variation ΔT_{cr} increased in proportion to R_{tp} and depended on the curvature of the CNTs, k. In particular, they reported that ΔT_{cr} varies approximately as $\Delta T_{cr} \propto k^{-2}$ and the increment rate, $\Delta T_{cr}/T_{cr-o}$, of the torsional buckling moment increases in a manner that is inversely proportional to CNT curvature.

Jeong et al. [35] also examined the effect of compressive prestress on the torsional responses of CNTs that have aspect ratios lower than about 12.5, below which compressive shell buckling occurs [17]. Figure 10-19 indicates that the variation rate $\Delta K_u/K_{u-o}$ in torsional stiffness by compressive prestress, σ_{cp}, decreases depending on only the compressive prestress ratio R_{cp} and not on CNT curvature. The torsional stiffness is decreased to approximately 21% at R_{cp}=0.82, which is where the minimum $\Delta K_u/K_{u-o}$ occurs. The $\Delta T_{cr}/T_{cr-o}$ depends only on R_{cp}. This implies that the CNT curvature dependence is different from what occurs in the case of the tensile prestress shown above. This can be explained using the stress concept of continuum

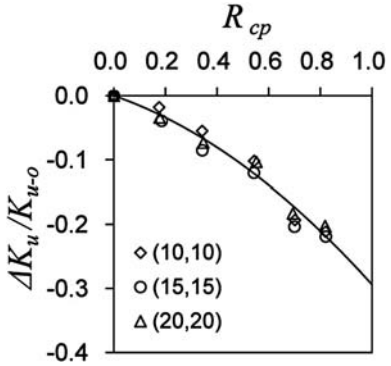


Figure 10-19. Variation rate $\Delta K_u/K_{u-o}$ of the torsional stiffness K_u by compressive prestress σ_{cp} in terms of compressive prestress ratio R_{cp}, which is the ratioσ_{cp}/σ_{cr} of compressive prestress and critical stress. After [35] with permission of AIP (2008)

mechanics theory, which tells us that the tensile prestress is the same for all CNTs at the same ratio of tensile prestress and strength because their tensile strength is constant [42]. However, the compressive prestress varies for different CNTs even if they have the same ratio of compressive prestress and critical stress because critical compressive stress σ_{cr} depends on CNT curvature [1].

As mentioned above, $\Delta K_u/K_{u-o}$ does not depend on CNT curvature. In addition, $\Delta K_u/K_{u-o}$ also does not depend on the length of the CNTs because it only expresses the net variation rate in torsional stiffness. Here, the fact that the torsional stiffness is changed by axial prestress indicates nonlinear effects due to mechanical coupling between torsional shear stress and axial prestress. This unusual behavior is thought to be caused mainly by the ability of the CNTs to endure relatively large axial elongation and thus the trends may not be significantly different in the cases of zig-zag and multi-walled CNTs.

The torsional stiffness of conventional linear elastic materials is not altered by the axial prestress in cases of closed cross sections with relatively high torsional rigidity that are very similar to those of CNTs [112, 113]. Therefore, the changes in torsional stiffness of CNTs by the axial prestress are unique relative to linear elastic materials. This is most important for applications such as torsional oscillators because the torsional stiffness is directly associated with the torsional oscillation frequency. Consequently, the use of axial prestress may allow tuning of the torsional stiffness and oscillation frequency of CNTs.

In applications such as bending oscillators, the axial stress has been used to tune their bending oscillation frequency [11]. This is a common approach as in tuning of the frequency of a guitar string. However, the concept that the axial stress can tune the torsional stiffness and oscillation frequency is unfamiliar to most engineering fields. Another interesting fact comes from the nonlinear mechanical effects due to the change in torsional stiffness. In pure torsion, CNTs behave as linear torsonal oscillators [105]. However, the use of axial prestress can cause nonlinear torsional

oscillations due to nonlinear mechanical effects. It is well known that nonlinear oscillators have hysteretic behavior and bistability in their frequency response, and thus possibly could be used for many applications including sensitive mass sensors and electronics [11, 117]. Finally, the potential use of axial prestress may be more important in graphene which is a being considered for use in electronics and NEMS [118]. In graphene, the effects of axial stress on torsional stiffness and oscillation frequency would be much larger due to its low torsional rigidity.

10.3.4.5. Crystalline Bundle

In many cases, CNTs are self-organized into crystalline bundles by van der Waals interactions. Such CNT bundles can be used for various applications, including use as fibers in nanotube composite materials [15], and therefore their mechanical responses should be investigated to optimize their use in these applications.

While numerous studies have examined the torsional responses of SWNTs and MWNTs, there is much that is still unknown regarding other aspects of the torsional responses of the CNT bundles. MD simulations have been also popularly used to investigate the torsional behavior of CNT bundles and the influence of the bundle environment as in the works of *Jeong* et al. [32, 33] and Qian et al. [119]. Jeong et al. [32, 33] examined the torsional responses, such as critical loads and stiffness, of bundles, and then compared with the responses of individual CNTs. Their simulations indicated that the torsional buckling moment and torsional stiffness of a single CNT in a bundle are higher than those of the individual CNTs and the twisted configuration of this bundle is similar to that of macroscopic ropes. In addition, they reported that the intertube van der Waals interactions within the bundle apparently increase its torsional buckling moment and torsional stiffness. Qian et al. [119] examined the nature of load transfer in a SWNT bundle consisting of one core tube surrounded by 6 tubes on the perimeter. They found that compared with parallel bundles, twisting the bundles can significantly enhance the load transfer between neighboring tubes in the bundles. This was computed from MD simulations as a function of twist angle for the nanotube bundles.

10.4. CONCLUSIONS

This chapter provides a focused discussion of the background of MD simulation methods, their role in the nanomechanics study of nanometer-scale CNTs, and their important contributions to this emerging research field. In particular, the illustrative examples present that show how these approaches are providing new and exciting insights into nanomechanics responsible for the mechanical behavior of CNTs, such as elastic moduli or stiffness, fracture or bucklng behavior, effect of filling or functionalization, effect of combined loads, effect of external gases, and related atomic-scale or molecular scale phenomena. The ability of MD simulations to study the nanomechanics of CNTs is revolutionizing our understanding of their mechanical properties and behavior at the most fundamental atomic level. In addition, these simulation methods have been also attracting a considerable interest in

the traditional fields of continuum-based mechanics as they can be applicable in investigating the microscopic mechanical behavior of materials [120]. Finally, methods such as MD simulations lead to excellent prospects for their contributions to nanomechanics or micromechanics study as they can simulate large-size systems and thus make possible the direct simulations with fully atomic-scale details with increasing computing power.

ACKNOWLEDGMENTS

We acknowledge support from the Network for Computational Nanotechnology through National Science Foundation Grant No. EEC-0228390.

REFERENCES

1. B. I. Yakobson, C. J. Brabec, and J. Bernholc, Physical Review Letters **76**, 2511 (1996).
2. H. Dai, J. H. Hafner, A. G. Rinzler, D. T. Colbert, and R. E. Smalley, Nature **384**, 147 (1996).
3. F. N. Dzegilenko, D. Srivastava, and S. Saini, Nanotechnology **10**, 253 (1999).
4. A. M. Fennimore, T. D. Yuzvinsky, W. Q. Han, M. S. Fuhrer, J. Cumings, and A. Zettl, Nature **424**, 408 (2003).
5. J. Kong, N. R. Franklin, C. W. Zhou, M. G. Chapline, S. Peng, K. J. Cho, and H. J. Dai, Science **287**, 622 (2000).
6. Y. Maniwa, K. Matsuda, H. Kyakuno, S. Ogasawara, T. Hibi, H. Kadowaki, S. Suzuki, Y. Achiba, and H. Kataura, Nature Materials **6**, 135 (2007).
7. J. C. Meyer, M. Paillet, and S. Roth, Science **309**, 1539 (2005).
8. G. Nagy, M. Levy, R. Scarmozzino, R. M. Osgood Jr, H. Dai, R. E. Smalley, C. A. Michaels, G. W. Flynn, and G. F. McLane, Applied Physics Letters **73**, 529 (1998).
9. S. J. Papadakis, A. R. Hall, P. A. Williams, L. Vicci, M. R. Falvo, R. Superfine, and S. Washburn, Physical Review Letters **93**, 146101 (2004).
10. P. Poncharal, Z. L. Wang, D. Ugarte, and W. A. de Heer, Science **283**, 1513 (1999).
11. V. Sazonova, Y. Yaish, H. Ustunel, D. Roundy, T. A. Arias and P. L. McEuen, Nature **431**, 284 (2004).
12. P. A. Williams, S. J. Papadakis, A. M. Patel, M. R. Falvo, S. Washburn, and R. Superfine, Applied Physics Letters **82**, 805 (2003).
13. S. S. Wong, E. Joselevich, A. T. Woolley, C. L. Cheung, and C. M. Lieber, Nature **394**, 52 (1998).
14. S. S. Wong, A. T. Woolley, T. W. Odom, J. L. Huang, P. Kim, D. V. Vezenov, and C. M. Lieber, Applied Physics Letters **73**, 3465 (1998).
15. R. H. Baughman, A. A. Zakhidov, and W. A. de Heer, Science **297**, 787 (2002).
16. M. Arroyo, and T. Belytschko, Physical Review Letters **91**, 215505 (2003).
17. M. J. Buehler, Y. Kong, and H. Gao, Journal of Engineering Materials and Technology **126**, 245 (2004).
18. G. Cao, and X. Chen, Physical Review B **73**, 155435 (2006).
19. K. Chandraseker, and S. Mukherjee, Journal of Applied Mechanics **73**, 315 (2006).
20. D. Danailov, P. Keblinski, S. Nayak, and P. M. Ajayan, Journal of Nanoscience and Nanotechnology **2**, 503 (2002).
21. M. Huhtala, A. V. Krasheninnikov, J. Aittoniemi, S. J. Stuart, K. Nordlund, and K. Kaski, Physical Review B **70**, 45404 (2004).
22. S. Iijima, C. Brabec, A. Maiti, and J. Bernholc, The Journal of Chemical Physics **104**, 2089 (1996).
23. A. Kutana, and K. P. Giapis, Physical Review Letters **97**, 245501 (2006).

24. K. M. Liew, C. H. Wong, and M. J. Tan, Applied Physics Letters **87**, 041901 (2005).
25. S. L. Mielke, D. Troya, S. Zhang, J. L. Li, S. Xiao, R. Car, R. S. Ruoff, G. C. Schatz, and T. Belytschko, Chemical Physics Letters **390**, 413 (2004).
26. T. Ozaki, Y. Iwasa, and T. Mitani, Physical Review Letters **84**, 1712 (2000).
27. L. Wang, and H. Hu, Physical Review B **71**, 195412 (2005).
28. P. Zhang, Y. Huang, P. H. Geubelle, P. A. Klein, and K. C. Hwang, International Journal of Solids and Structures **39**, 3893 (2002).
29. S. L. Zhang, S. L. Mielke, R. Khare, D. Troya, R. S. Ruoff, G. C. Schatz, and T. Belytschko, Physical Review B **71**, 115403 (2005).
30. S. J. Heo, and S. B. Sinnott, Journal of Nanoscience and Nanotechnology **7**, 1518 (2007).
31. B. W. Jeong, and J. K. Lim, Transactions of the KSME (A) **30**, 809 (2006).
32. B. W. Jeong, and J. K. Lim, Transactions of the KSME (A) **31**, 862 (2007).
33. B. W. Jeong, J. K. Lim, and S. B. Sinnott, Journal of Applied Physics **101**, 084309 (2007).
34. B. W. Jeong, J. K. Lim, and S. B. Sinnott, Applied Physics Letters **91**, 093102 (2007).
35. B. W. Jeong, J. K. Lim, and S. B. Sinnott, Applied Physics Letters **92**, 253114 (2008).
36. B. Ni, S. B. Sinnott, P. T. Mikulski, and J. A. Harrison, Physical Review Letters **88**, 205505 (2002).
37. S. B. Sinnott, O. A. Shenderova, C. T. White, and D. W. Brenner, Carbon **36**, 1 (1998).
38. T. Belytschko, S. P. Xiao, G. C. Schatz, and R. S. Ruoff, Physical Review B **65**, 235430 (2002).
39. B. W. Jeong, J. K. Lim, and S. B. Sinnott, Applied Physics Letters **90**, 023102 (2007).
40. B. W. Jeong, J. K. Lim, and S. B. Sinnott, Nanotechnology **18**, 485715 (2007).
41. M. A. L. Marques, H. E. Troiani, M. Miki-Yoshida, M. Jose-Yacaman, and A. Rubio, Nano Letters **4**, 811 (2004).
42. M. Sammalkorpi, A. Krasheninnikov, A. Kuronen, K. Nordlund, and K. Kaski, Physical Review B **70**, 245416 (2004).
43. D. Srivastava, C. Wei, and K. Cho, Applied Mechanics Reviews **56**, 215 (2003).
44. C. Wei, K. Cho, and D. Srivastava, Physical Review B **67**, 115407 (2003).
45. B. I. Yakobson, M. P. Campbell, C. J. Brabec, and J. Bernholc, Computational Materials Science **8**, 341 (1997).
46. L. G. Zhou, and S. Q. Shi, Computational Materials Science **23**, 166 (2002).
47. A. Garg, J. Han, and S. B. Sinnott, Physical Review Letters **81**, 2260 (1998).
48. A. Garg, and S. B. Sinnott, Physical Review B **60**, 13786 (1999).
49. A. Garg, and S. B. Sinnott, Chemical Physics Letters **295**, 273 (1998).
50. Z. G. Mao, A. Garg, and S. B. Sinnott, Nanotechnology **10**, 273 (1999).
51. H. Trotter, R. Phillips, B. Ni, Y. H. Hu, S. B. Sinnott, P. T. Mikulski, and J. A. Harrison, Journal of Nanoscience and Nanotechnology **5**, 536 (2005).
52. K. H. Lee, P. Keblinski, and S. B. Sinnott, Nano Letters **5**, 263 (2005).
53. S. J. Heo, and S. B. Sinnott, Journal of Applied Physics **102**, 064307 (2007).
54. I. Jang, S. B. Sinnott, D. Danailov, and P. Keblinski, Nano Letters **4**, 109 (2004).
55. B. Ni, and S. B. Sinnott, Surface Science **487**, 87 (2001).
56. S. K. Pregler, B. W. Jeong, and S. B. Sinnott, Composites Science and Technology **68**, 2049 (2008).
57. S. K. Pregler, and S. B. Sinnott, Physical Review B **73**, 224106 (2006).
58. W. G. Sawyer, S. S. Perry, S. R. Phillpot, and S. B. Sinnott, Journal of Physics Condensed Matter **20**, 354012 (2008).
59. B. J. Alder, and T. E. Wainwright, The Journal of Chemical Physics **27**, 1208 (2004).
60. B. J. Alder, and T. E. Wainwright, The Journal of Chemical Physics **31**, 459 (2004).
61. M. P. Allen, and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
62. M. Finnis, *Interatomic Forces in Condensed Matter* (Oxford University Press, Oxford, 2003).
63. D. Frenkel, and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, New York, 2002).

64. J. M. Haile, *Molecular Dynamics Simulation: Elementary Methods* (Wiley, New York, 1992).
65. G. C. Abell, Physical Review B **31**, 6184 (1985).
66. J. Tersoff, Physical Review Letters **56**, 632 (1986).
67. J. Tersoff, Physical Review Letters **61**, 2879 (1988).
68. D. W. Brenner, Physical Review B **42**, 9458 (1990).
69. D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott, Journal of Physics Condensed Matter **14**, 783 (2002).
70. F. H. Stillinger, and T. A. Weber, Physical Review B **31**, 5262 (1985).
71. M. S. Daw, and M. I. Baskes, Physical Review Letters **50**, 1285 (1983).
72. S. M. Foiles, Physical Review B **32**, 3409 (1985).
73. M. I. Baskes, Physical Review B **46**, 2727 (1992).
74. M. I. Baskes, J. S. Nelson, and A. F. Wright, Physical Review B **40**, 6085 (1989).
75. K. Ohno, K. Esfarjani, and Y. Kawazoe, *Computational Materials Science: From Ab Initio to Monte Carlo Methods* (Springer, Berlin, 1999).
76. A. K. Rappe, and W. A. Goddard III, The Journal of Physical Chemistry **95**, 3358 (1991).
77. S. J. Stuart, A. B. Tutein, and J. A. Harrison, The Journal of Chemical Physics **112**, 6472 (2000).
78. J. E. Lennard-Jones, Proceedings of the Physical Society **43**, 461 (1931).
79. H. C. Andersen, The Journal of Chemical Physics **72**, 2384 (1980).
80. S. Nose, The Journal of Chemical Physics **81**, 511 (1984).
81. W. G. Hoover, Physical Review A **31**, 1695 (1985).
82. J. E. Shigley, and C. R. Mischke, *Mechanical Engineering Design* (McGraw Hill Publishing Company, Singapore, 2001).
83. A. C. Ugural, and S. K. Fenster, *Advanced Strength And Applied Elasticity* (Prentice-Hall Publishing, Englewood Cliffs, NJ, 2003).
84. J. P. Lu, Physical Review Letters **79**, 1297 (1997).
85. B. WenXing, Z. ChangChun, and C. WanZhao, Physica B: Physics of Condensed Matter **352**, 156 (2004).
86. D. H. Robertson, D. W. Brenner, and J. W. Mintmire, Physical Review B **45**, 12592 (1992).
87. E. Hernandez, C. Goze, P. Bernier, and A. Rubio, Physical Review Letters **80**, 4502 (1998).
88. M. M. J. Treacy, T. W. Ebbesen, and J. M. Gibson, Nature **381**, 678 (1996).
89. M. F. Yu, B. S. Files, S. Arepalli, and R. S. Ruoff, Physical Review Letters **84**, 5552 (2000).
90. M. F. Yu, O. Lourie, M. J. Dyer, K. Moloni, T. F. Kelly, and R. S. Ruoff, Science **287**, 637 (2000).
91. B. W. Smith, and D. E. Luzzi, Chemical Physics Letters **321**, 169 (2000).
92. R. K. Saini, I. W. Chiang, H. Q. Peng, R. E. Smalley, W. E. Billups, R. H. Hauge, and J. L. Margrave, Journal of the American Chemical Society **125**, 3617 (2003).
93. P. A. Williams, S. J. Papadakis, A. M. Patel, M. R. Falvo, S. Washburn, and R. Superfine, Physical Review Letters **89**, 255502 (2002).
94. P. C. Andia, F. Costanzo, and G. L. Gray, Modelling and Simulation in Materials Science and Engineering **14**, 741 (2006).
95. K. S. Cheung, and S. Yip, Journal of Applied Physics **70**, 5688 (1991).
96. D. H. Tsai, The Journal of Chemical Physics **70**, 1375 (1979).
97. M. Zhou, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **459**, 2347 (2003).
98. C. F. Cornwell, and L. T. Wille, The Journal of Chemical Physics **109**, 763 (1998).
99. H. Dai, N. Franklin, and J. Han, Applied Physics Letters **73**, 1508 (1998).
100. J. A. Harrison, S. J. Stuart, D. H. Robertson, and C. T. White, Journal of Physical Chemistry B **101**, 9682 (1997).
101. S. P. Timoshenko, and J. M. Gere, *Theory of Elastic Stability* (McGraw Hill, New York, 1961).
102. S.-J. Heo, and S. B. Sinnott, (in review).

103. J. Chen, M. A. Hamon, H. Hu, Y. Chen, A. M. Rao, P. C. Eklund, and R. C. Haddon, Science **282**, 95 (1998).
104. S. J. Heo, and S. B. Sinnott, (in review).
105. T. Cohen-Karni, L. Segev, O. Srur-Lavi, S. R. Cohen, and E. Joselevich, Nature Nanotechnology **1**, 36 (2006).
106. A. R. Hall, L. An, J. Liu, L. Vicci, M. R. Falvo, R. Superfine, and S. Washburn, Physical Review Letters **96**, 256102 (2006).
107. Y. M. Ying, R. K. Saini, F. Liang, A. K. Sadana, and W. E. Billups, Organic Letters **5**, 1471 (2003).
108. C. Q. Ru, Journal of the Mechanics and Physics of Solids **49**, 1265 (2001).
109. B. Popescu, and D. H. Hodges, International Journal of Non-Linear Mechanics **34**, 709 (1999).
110. M. K. Hurst, *Prestressed Concrete Design* (Taylor & Francis, New York, 1998).
111. D. Zhou, M. Kamlah, and D. Munz, Journal of the European Ceramic Society **25**, 425 (2005).
112. J. A. Connally, *Torsion of a Thin Rectangular Beam with Axial Prestress and Ends Constrained From Warping* (Massachusetts Institute of Technology, Cambridge, MA, 1986).
113. M. A. Biot, Journal of Applied Physics **10**, 860 (1939).
114. H. Kolsky, and A. C. Pipkin, Archive of Applied Mechanics (Ingenieur Archiv) **49**, 337 (1980).
115. N. Wang, I. M. Tolic-Norrelykke, J. Chen, S. M. Mijailovich, J. P. Butler, J. J. Fredberg, and D. Stamenovic, American Journal of Physiology. Cell Physiology **282**, 606 (2002).
116. E. Joselevich, Chemphyschem : A European Journal of Chemical Physics and Physical Chemistry **7**, 1405 (2006).
117. H. B. Chan, V. A. Aksyuk, R. N. Kleiman, D. J. Bishop, and F. Capasso, Physical Review Letters **87**, 211801 (2001).
118. A. K. Geim, and K. S. Novoselov, Nature Materials **6**, 183 (2007).
119. D. Qian, W. K. Liu, and R. S. Ruoff, Composites Science and Technology **63**, 1561 (2003).
120. W. K. Liu, S. Jun, and D. Qian, Journal of Computational and Theoretical Nanoscience **5**, 970 (2008).

CHAPTER 11

STICK-SPIRAL MODEL FOR STUDYING MECHANICAL PROPERTIES OF CARBON NANOTUBES

TIENCHONG CHANG

Shanghai Institute of Applied Mathematics and Mechanics, Institute of Low Dimensional Carbon and Device Physics, Shanghai University, Shanghai 200072, People's Republic of China,
e-mail: tchang@staff.shu.edu.cn

Abstract: Quantum/molecular mechanics and continuum mechanics have been highly developed to describe material properties at small and large length scales. As we enter the era of nanotechnology, it has become increasingly important to model phenomena at mesoscopic length scales. Two alternative approaches, namely the "bottom up" approach based on quantum/molecular mechanics and the "top down" approach based on continuum mechanics, are frequently used to model mechanical properties of nano-structured materials. However, the connection between these two approaches is not well established. Much effort has been made to develop theories and approaches to span multiple length scales or to bridge gap between the two approaches. Based on a molecular mechanics concept, a stick-spiral model is developed to analytically link the molecular structure and macroscopic properties of carbon nanotubes. We review and summarize in this chapter the recent advances on this model

Keywords: Carbon nanotubes, Mechanical properties, Multiscale model

11.1. INTRODUCTION

Describing material behavior is a key issue in engineering applications. At different length scales, the same material may behave much differently due to its hierarchical structure. Different theories have been developed to describe mechanical behaviors of materials at different length scales. For example, quantum mechanics predicts mechanical behavior of materials based on calculations of electronic structures of molecules; molecular mechanics/dynamics ignores the electronic structure and predicts mechanical behavior of materials by calculating the system energy that is a function of the nuclear positions; while continuum mechanics, which ignores the discrete nature of atomic structure and treats materials as continuum, predicts material behavior by solving a series of analytical equations associated with equilibrium, compatibility and constitution. However, gaps remain between these theories and much effort has been made to establish organic connections [1–11].

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 297–322.
DOI 10.1007/978-1-4020-9785-0_11,

Advances in the synthesis of nanoscale materials have stimulated ever-broader research activities due to the combination of their expected structural perfection, small size, and excellent mechanical and electronic properties. However, a common theory to describe mechanical behavior of nanomaterials at this length scale is to date not well established. Two alternative approaches, namely the "bottom up" approach based on quantum/molecular mechanics and the "top down" approach based on continuum mechanics, are frequently used. Bottom-up approaches may, in principle, be used to investigate the behavior of any systems if atomic interactions could be determined. However, direct simulation of large scale problems (with long time or large volume) remains a heavy computer task. "Top down" approaches based on continuum mechanics is computationally efficient, but difficult to account for some effects such as those resulting from discrete atomic structures. Multi-scale simulations that bridge bottom up and top down approaches have been widely developed, such as hierarchical methods, concurrent methods, and quasicontinuum methods. A detailed review on this topic can be found in Liu et al. [6].

Carbon nanotubes (CNTs) are one of the most promising nanomaterials. The amazing mechanical properties such as exceptional high stiffness and tensile strength of CNTs make them highly potential and ideal candidates for multifarious applications including super-strong materials and nanomechanical devices [12, 13]. Efforts have been made to investigate the mechanical properties of CNTs because a clear understanding of these properties is essential to ensure the optimum performance of CNTs in potential applications. Based on a molecular mechanics concept, a stick-spiral model is developed to analytically link the molecular structure and macroscopic properties of carbon nanotubes [5, 14–17]. We review and summarize in this chapter the recent advances on this model.

The organization of this chapter is as follows. Section 11.2 briefly introduces CNTs and their mechanical properties, focusing in particular the theoretical modeling on geometry dependent mechanical properties of CNTs. Section 11.3 gives the description of the stick-spiral model and the formulation of its governing equations, and then presents linear and nonlinear stick-spiral models with their applications in predicting elastic properties and mechanical behaviors of CNTs. Section 11.4 offers concluding remarks.

11.2. CARBON NANOTUBES AND THEIR MECHANICAL PROPERTIES

11.2.1. Carbon Nanotubes (CNTs)

Carbon nanotubes, discovered by Iijima and coworkers [18, 19], can be classified into multi-walled carbon nanotubes (MWCNTs) and single-walled carbon nanotubes (SWCNTs). A MWCNT consists of two or more concentric cylindrical shells of graphene sheets arranged coaxially around a central hollow with interlayer separation as in graphite (0.34 nm), whereas a SWCNT is made of a single layer of graphite.

A SWCNT can be viewed as a graphene sheet rolled into a seamless tube. In principle, an infinite number of nanotube geometries can exist because a graphene sheet can be rolled up with different angles. Different rolling angles result in different chiralities, or helicities, of SWCNTs. A common approach is using a chiral vector or chiral angle to identify a SWCNT. Figure 11-1 shows a schematic illustration of a graphene sheet. A vector **C** in the graphene plane can be described as a combination of base vectors **a** and **b** of the hexagon by

$$\mathbf{C} = n\mathbf{a} + m\mathbf{b}, \tag{11-1}$$

with n and m being two integers. If the head of the vector **C** touches its tail when the graphene sheet rolled into a tube, we call **C** the chiral vector, or roll-up vector of the nanotube. The magnitude of the chiral vector, $C = \sqrt{3}r_0\sqrt{m^2 + n^2 + mn}$, represents the circumference of the nanotube, where r_0 is the carbon-carbon bond length. A SWCNT can thus be uniquely indexed by a pair of integers (n, m) [20] to represent its chirality or helicity. The chirality of a SWCNT can also be indicated by the chiral angle ϕ (see Figure 11-1) which is give by

$$\phi = \arccos \frac{2n + m}{2\sqrt{m^2 + n^2 + mn}}. \tag{11-2}$$

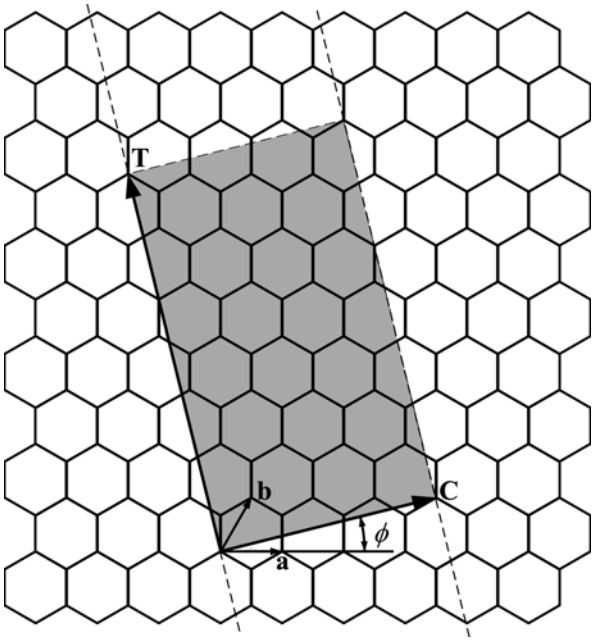


Figure 11-1. Schematic illustration of a graphene sheet and definitions of geometrical parameters used to describe a nanotube. A (3, 1) tube would be formed by rolling up the graphene sheet bounded by the *two dashed lines*. The unit cell of the nanotube is shown in *gray*

The two limiting cases of nanotubes are $(n, 0)$ (whose chiral angle is 0) and (n, n) (whose chiral angle is $\pi/6$), which are usually known as zigzag and armchair tubes based on the geometry of carbon bonds around the circumference of the nanotube. Zigzag and armchair tubes are achiral nanotubes whereas SWCNTs with a chiral angle of $0 < \phi < \pi/6$ are chiral nanotubes.

Another important geometrical parameter of SWCNTs is the translation vector **T**, which is directed along the SWCNT axis and perpendicular to the chiral vector **C** (see Figure 11-1). In the graphene plane, **T** is given by

$$\mathbf{T} = \frac{2m+n}{d_{\mathrm{R}}}\mathbf{a} - \frac{2n+m}{d_{\mathrm{R}}}\mathbf{b} \tag{11-3}$$

The magnitude of the translation vector, $T = 3r_0\sqrt{m^2+n^2+mn}/d_{\mathrm{R}}$, corresponds to the length of the SWCNT unit cell (which is marked in gray in Figure 11-1).

11.2.2. Mechanical Properties of CNTs

There are many experimental studies that gave direct proof of the exceptional mechanical properties of CNTs. By investigating vibration frequencies of cantilevered CNTs within a transmission electron microscope (TEM), Treacy et al. [21] obtained the Young's modulus of multi-walled carbon nanotubes (MWCNTs) in the range from 0.4 to 4.15 TPa, with a mean value of 1.8 TPa; Krishnan et al. [22] found the Young's modulus of single-walled carbon nanotubes (SWCNTs) varying from 0.9 to 1.7 TPa; Poncharal et al. [23] reported the Young's modulus of about 1 TPa for small diameter MWCNTs, while for the MWCNT with large diameters, the Young's modulus would be dramatically reduced up to 1 order because of the presence of rippling. By analyzing the bending behavior of MWCNTs which were manipulated by an atomic force microscope (AFM), Wong et al. [24] obtained an average value for the Young's modulus of 1.28 ± 0.59 TPa and Salvetat et al. [25] obtained a value of $0.81^{+0.41}_{-0.16}$ TPa, with no distinct dependence on the tube diameter. Yu et al. [26] directly applied axial tensile force on both ends of a MWCNT using two AFM tips, and measured the Young modulus ranging from 0.27 to 0.97 TPa and the tensile strength of the outmost layer varying from 11 to 63 GPa. A unique failure mode of the MWCNT ("sword-in-sheath") was firstly observed in the work. Another direct measurement performed by Demczyk et al. [27] indicated that the tensile strength of a MWCNT is about 150 GPa, and a value of 0.97 TPa derived for Young's modulus from a bending test was also reported.

11.2.3. Theoretical Modeling on Geometry Dependent Mechanical Properties of CNTs

Theoretical studies may provide more detailed information than an experimental investigation because a simultaneous measure of both mechanical properties and structural details (such as the chirality) of a CNT remains, to date, a challenge. Two categories of theoretical approaches, namely the bottom-up approach based

on quantum/molecular mechanics and the top-down approach based on continuum mechanics, are frequently used to study mechanical properties of nano-structured materials. Most of bottom-up approaches need a numerical procedure, such as molecular dynamics simulations, while many top-down methods are capable of giving analytical solutions to the problems considered.

Elastic properties and mechanical behavior of CNTs have been extensively studied by bottom-up calculations [28–48], of which some studies paid special concerns on the effect of the structural details. Based on the Tersoff-Brenner potential, Robertson et al. [28] predicted by molecular dynamics that the elastic constants along the tube axis generally soften with decreasing tube radius. Similar results were reported in tight binding calculations by Hernandez et al. [32], ab initio calculations by Sanchez-Portal et al. [34], and lattice dynamics calculations by Popov et al. [35]. Some of these works showed in particular that the mechanical behavior of a SWCNT is dependent on the tube chirality. The effects of geometrical detail on the tensile strength (or failure strain) [49–52, 36, 53, 38, 54, 45, 55] and the buckling behavior [38, 40, 56, 41, 46–48] of a SWCNT have also been investigated recently. Two types of failure process, i.e., brittle fracture due to direct bond breaking and plastic deformation due to dislocation (Stone-Wales defect) evolution, were studied in these studies [49–54, 45, 55]. Nardelli and coworkers [50, 66, 52] 2002 showed by ab initio calculations that the mechanical behavior (including strain release mechanism and failure process) of nanotubes under large tensile strain strongly depends on their chirality and diameter. Extensive discussions on these dependences can also be found in a recent paper by Dumitrica et al. [45]. In particular, they observed by quantum mechanical calculations that the failure strain of a SWCNT increases from $\sim$15% to $\sim$22% with the tube chiral angle increasing from 0 to 30°. Molecular dynamics simulations showed that buckling behavior of a SWCNT upon axial compression [57] and torsion [48] is dependent on the tube chirality. Molecular dynamics simulations indicate also that the initial buckling modes for thin and thick MWCNTs are quite different [46], which is confirmed by successive experimental observations [58]. Although bottom up calculations have been important tools in studying mechanical behavior of CNTs, direct simulation of large scale problems (with long time or large number of atoms) remains a heavy computer task. To reduce computational costs, many novel numerical approaches [59–72] have been proposed to investigate the mechanical behavior of CNTs, which makes some large scale problems such as buckling of thick MWCNTs capable of being studied. Besides, much effort has been made to derive explicit solutions based on continuum mechanics. For instance, thin shell theory was frequently use to analytically study buckling of CNTs [30, 73–84]. The applicability and limitations of shell models was outlined in Wang et al. [85]. Compared to the numerical methods, a shell model may give explicit solutions to the problems considered. However, at least two issues are not well addressed in using continuum mechanics models. First, some parameters of CNTs are not well defined in the continuum mechanics regime. For instance, no unique definition for the CNT wall thickness is appropriate for all purposes and conceptual ambiguities remain [e.g., Yakobson's paradox [30]]. To bypass problems resulted from ill-defined tube wall thickness, some effective

parameters, such as effective bending stiffness [86], and effective elastic modulus [32] were proposed. Second, and more importantly, it is difficult to account for the possible effect of tube chirality on the behavior of CNTs in a continuum mechanics model because in which the discrete nature of atomic structures is erased.

Recently, explicit solutions for the elastic properties of CNTs were derived within the framework of molecular mechanics, making the mechanical behavior of CNTs possible to be analytically investigated by bottom-up approaches. A truss model is presented by Odegard et al. [87] to establish the relation between effective bending rigidity and molecular properties of a graphene sheet by equating the molecular potential energy to the mechanical strain energy. A similar model was used by Wang [88] to obtain the effective in-plane stiffness and bending rigidity of achiral (i.e., armchair and zigzag) SWCNTs. Chang and Gao [5] established a "stick-spiral" model and obtained the first closed-form expressions for the longitudinal Young's modulus and Poisson's ratio of achiral SWCNTs. The closed-form expressions for elastic properties of achiral SWCNTs under various loading conditions were presented by Shen and Li [89] via a energy approach. Leung et al. [90] obtained Young's modulus for zigzag CNTs by developing an equivalent truss model. The closed-form expressions for the axial elastic properties of chiral carbon nanotubes were presented recently by [14, 16]. These expressions are concise but capable of directly linking material properties at different length scales. The effects of structural details on the elastic properties of CNTs can thus be reflected in these expressions. The stick-spiral model can also be used to investigate the effect of structural details on the buckling of CNTs [91–93]. The model in the above-mentioned studies is indeed a linear model because in which harmonic potentials are used to model interatomic interactions. To study mechanical behavior of CNTs under large strains, a nonlinear model must be developed. Xiao et al. [94] 2006 firstly extended Chang and Gao's [5] model to investigate nonlinear behavior of achiral SWCNTs by incorporating a Morse type potential presented by Belytschko et al. [36]. Geng and Chang [15] further extended the model to evaluate nonlinear mechanical behavior of SWCNTs with arbitrary chirality. Most recently, Duan et al. [95] developed a nonlinear stick-spiral model by incorporating a reactive empirical bond-order potential given by Brenner et al. [96], and investigated the fracture behavior of SWCNTs. The analytical results from the mentioned molecular mechanics models for the elastic properties and mechanical behavior are in reasonable agreement with the existing numerical results, but one needs much less computer time to yield them.

11.3. STICK-SPIRAL MODEL FOR CARBON NANOTUBES

11.3.1. Model Description

In the empirical force field method of molecular mechanics, the total potential energy, E_t, can be expressed as a sum of several individual energy contributions

$$E_t = U_r + U_\theta + U_\tau + U_\omega + U_{vdW} + U_{es}, \tag{11-4}$$

where bonded terms U_r, U_θ, U_ω, and U_τ are energies associated with bond stretching, bond angle variation, bond inversion, and torsion, respectively; and nonbonded terms U_{vdW} and U_{es} are associated with van der Waals and electrostatic interactions [97, 98, 5], respectively. Various functional forms may be used for these energy terms depending on the particular materials and loading conditions considered.

In some cases, some of the energy terms in the right hand side of Eq. (11-1) may be ignored. For example, if all atom of a SWCNT is kept in cylindrical surfaces when it deforms, it can be expected that only bond stretching (U_r) and bond angle variation (U_θ) terms are significant in the total system potential energy.

A SWCNT, which can be viewed as a graphene sheet rolled into a seamless tube, is usually indexed by a pair of integers (n, m) to represent its helicity [20]. We consider a (n, m) tube subjected to an axial force F, an internal pressure P, and an axial torque M_{T}, as shown in Figure 11-2a. There are three chemical bonds r_1, r_2, r_3 and three bond angles θ_1, θ_2, θ_3 associated with each carbon atom (Figure 11-2b). The relationship between external stresses and variations in bond lengths and bond angles can be determined via equilibrium and geometry of the structure.

To obtain the equilibrium equations of the structure, we can visualize the present molecular mechanical model as an effective "stick-spiral" system. In this system, we use a stick with an axial stiffness of $K_r(r) = \partial^2 U_r / \partial r^2$ (in which r is the bond

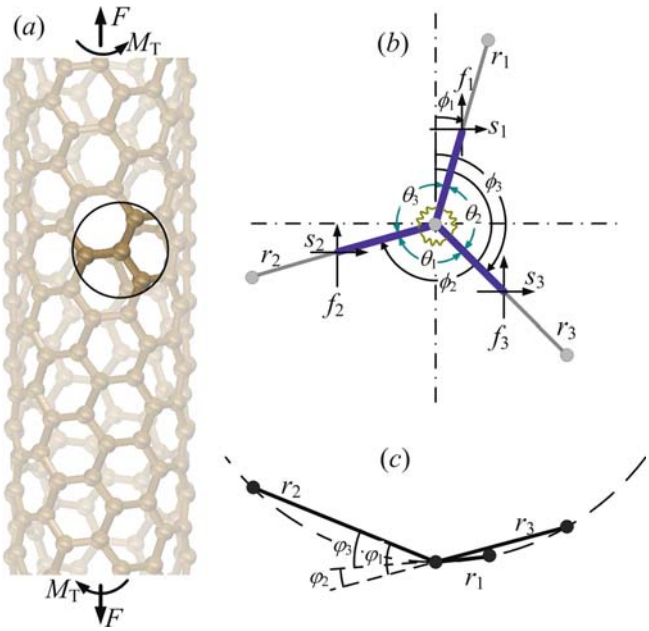


Figure 11-2. (**a**) Global structure of a chiral nanotube. (**b**) Front view of local structure and schematic of the stick-spiral model. In the model, a stick with infinite bending stiffness is used to model force-stretch relationship of the carbon–carbon bond, and a spiral spring is used to describe the twisting moment resulting from an angular distortion of the bond angle. (**c**) Top view of local structure

length) to model the stretching response of the carbon–carbon bond and a spiral spring with a stiffness of $K_\theta(\theta) = \partial^2 U_\theta / \partial\theta^2$ (in which θ is the bond angle) to model the twisting response resulting from an angular distortion of bond angle. In addition, we assume that the stick has an infinite bending stiffness because the chemical bond always remains straight irrespective of the applied load. This provides a clear picture of the mechanical model of the problem in hand. In the stick-spiral model, the stretching force resulting from bond elongation and the twisting moment resulting from bond angle variation is calculated by

$$F^*(\Delta r) = \int_0^{\Delta r} K_r(r_0 + \delta)\mathrm{d}\delta = \partial U_r / \partial r|_{r_0+\Delta r}\,, \tag{11-5}$$

$$M^*(\Delta\theta) = \int_0^{\Delta\theta} K_\theta(\theta_0 + \delta)\mathrm{d}\delta = \partial U_\theta / \partial\theta|_{\theta_0+\Delta\theta}\,, \tag{11-6}$$

where r_0 and Δr, and θ_0 and $\Delta\theta$ are the equilibrium bond length and the elongation of bond length, and the equilibrium bond angle and the variation of bond angle, respectively.

11.3.2. Governing Equations of the Stick-Spiral Model

The relationships between the external forces and the internal forces yield

$$F = (n + m)f_1 + mf_2, \tag{11-7}$$

$$P = -[(n - m)s_1 + (2n + m)s_2]/(TR), \tag{11-8}$$

$$M_\mathrm{T} = [(n + m)s_1 + ms_2]R, \tag{11-9}$$

where R is the tube radius, and f_i and s_i are forces contributed on carbon bond along axial and circumferential directions, respectively (see Figure 11-2b).

Equilibrium of the local structure of the SWCNT needs

$$f_1 + f_2 + f_3 = 0, \tag{11-10}$$

$$s_1 + s_2 + s_3 = 0. \tag{11-11}$$

We now consider the forces and moment in stick (bond) r_i with an orientation angle (the angle between the bond and the tube axis, see Figure 11-2) of ϕ_i. Decompose the force f_i into two components, of which one is $f_i\cos\phi_i$ along the bond direction, and the other $f_i\sin\phi_i$ perpendicular to the bond. The force s_i can be decomposed similarly, but it should be noted that $s_i\sin\phi_i$ is along the bond direction while $s_i\cos\phi_i$ is perpendicular to the bond.

Force equilibrium to bond extension leads to

$$f_i \cos\phi_i + s_i \sin\phi_i = K_\rho \mathrm{d}r_i \quad i = 1,2,3. \tag{11-12}$$

The internal bending moment distribution in a stick r_i is antisymmetric about its midpoint. This means that the internal bending moment at the midpoint of the stick r_i is zero. It is thus convenient to establish the moment equilibrium equation by dividing the stick r_i into two equal halves. Let us focus attention on one half of the stick r_i. The effective lateral force exerted on this considered half by the other half is $f_i\sin\phi_i$-$s_i\cos\phi_i$, which exerts a rotation moment of $(f_i\sin\phi_i$-$s_i\cos\phi_i)r_i/2$ on the considered half. On the other hand, rotation moments produced by bond angle variance $\Delta\theta_j$ is $\partial U_{\theta_j}/\partial\theta_j$, and by $\Delta\theta_k$ is $\partial U_{\theta_k}/\partial\theta_k$. The moment equilibrium of the considered half can thus be expressed by

$$\frac{r_i}{2}(f_i\sin\phi_i - s_i\cos\phi_i) = \frac{\partial U_{\theta_j}}{\partial\theta_j}\cos\omega_{ij} + \frac{\partial U_{\theta_k}}{\partial\theta_k}\cos\omega_{ik} \quad i,j,k = 1,2,3; i \neq j \neq k, \tag{11-13}$$

where ω_{ij}, the torsion angle between the plane though r_i parallel to the nanotube axis and the plane of θ_j, can be calculated by (see Appendix)

$$\cos\omega_{ij} = (\cos\phi_i\sin\phi_k\cos\varphi_j - \sin\phi_i\cos\phi_k)/\sin\theta_j \quad i,j,k = 1,2,3; i \neq j \neq k. \tag{11-14}$$

The structural parameters φ_i, ϕ_i and θ_i are defined in Figure 11-2.
The geometrical relationships of a SWCNT satisfy (see Appendix)

$$\cos\theta_i = \sin\phi_j\sin\phi_k\cos\varphi_i + \cos\phi_j\cos\phi_k \quad i,j,k = 1,2,3; i \neq j \neq k. \tag{11-15}$$

With use of Eqs. (11-14) and (11-15), the variation of bond angle can be obtained as

$$\Delta\theta_i = -(\Delta\phi_j\cos\omega_{ji} + \Delta\phi_k\cos\omega_{ki}) \quad i,j,k = 1,2,3; i \neq j \neq k. \tag{11-16}$$

Cylindrical structure of a defect-free SWCNT always needs its chiral vector to keep a closed ring, i.e., the dislocation between the head and the tail of the chiral vector should be zero no matter how the SWCNT was deformed unless the presence of defects. This feature actually gives compatible equation of a deformed SWCNT as follows

$$\Delta[mr_1\cos\phi_1 - (n+m)r_2\cos\phi_2 + nr_3\cos\phi_3] = 0. \tag{11-17}$$

We now have 15 independent equations given by Eqs. (11-7)–(11-13), (11-16), and (11-17), and 18 independent variables F, P, M_T, f_1, f_2, f_3, s_1, s_2, s_3, $\mathrm{d}r_1$, $\mathrm{d}r_2$, $\mathrm{d}r_3$, $\mathrm{d}\theta_1$, $\mathrm{d}\theta_2$, $\mathrm{d}\theta_3$, $\mathrm{d}\phi_1$, $\mathrm{d}\phi_2$, $\mathrm{d}\phi_3$. We note here that F, P, and M_T are applied external forces. Once these forces are given, the problem is solvable.

11.3.3. Linear Stick-Spiral Model and its Applications

11.3.3.1. Linear Stick-Spiral Model

There are many functional forms to model the molecular potential. A complex function must be used to describe the behavior of a chemical bond far from its equilibrium position. An example is the Morse potential in which three parameters is used to describe the behavior from equilibrium to bond dissociation. For most problems, harmonic functions suffice to characterize atomic interactions under relatively small strain conditions. Chemical calculations have shown that harmonic functions provide a reasonable approximation to the potential energy of molecular systems in which the bond length is near its equilibrium. When harmonic potentials are used, the total system potential of the stick-spiral model can be expressed as

$$E = U_r + U_\theta = \frac{1}{2}\sum_i K_r(\Delta r_i)^2 + \frac{1}{2}\sum_j K_\theta(\Delta\theta_j)^2, \tag{11-18}$$

while the stretching force resulting from bond elongation and the twisting moment resulting from bond angle variation are calculated by

$$F^*(\Delta r_i) = K_r \Delta r_i, \tag{11-19}$$

$$M^*(\Delta\theta_j) = K_\theta \Delta\theta_j. \tag{11-20}$$

This is in fact a linear stick-spiral model since the stick has a constant stiffness K_r and the spiral spring has a constant stiffness K_θ.

11.3.3.2. Elastic Mechanical Properties of SWCNTs

Closed-form expressions for mechanical properties of a SWCNT can be obtained via the linear stick-spiral model. Here we present some of the expressions. Detailed procedure to get these expressions can be found in our previous works [5, 14, 16, 17].

The surface Young's modulus Y_S and the Poisson's ratio ν in axial direction are find to be [14, 16]

$$Y_S = \eta\frac{4\mu K_\rho}{\sqrt{3}(\lambda + 3\mu)}, \tag{11-21}$$

$$\nu = \frac{\lambda - \xi\mu}{\lambda + 3\mu}, \tag{11-22}$$

where η, λ, and ξ are parameters associated with n and m, and $\mu = K_\theta/K_r r_0^2$ with r_0 the equilibrium carbon-carbon bond length. In particular, we can obtain the values of these parameters for some limit cases as follows

$$\lambda = \frac{7 - \cos(\pi/n)}{34 + 2\cos(\pi/n)}, \eta = 1, \xi = 1, \text{ for armchair SWCNTs}(m = n), \tag{11-23}$$

$$\lambda = \frac{5 - 3\cos(\pi/n)}{14 - 2\cos(\pi/n)}, \eta = 1, \xi = 1, \text{ for zigzag SWCNTs}(m = 0), \tag{11-24}$$

$$\lambda = 1/6, \eta = 1, \xi = 1, \text{ for graphene sheets}(m, n \sim \infty). \tag{11-25}$$

The surface Young's modulus and Poisson's ratio in circumferential direction are found to be the same as those in axial direction [16].

The surface shear modulus G_S is obtained as [16]

$$G_S = \frac{\mu K_\rho}{\sqrt{3}(\bar{\lambda} + \bar{\zeta}\mu)}, \tag{11-26}$$

where $\bar{\lambda}$ and $\bar{\zeta}$ are parameters associated with n and m. The values of the two parameters for some limit cases are as follows

$$\bar{\lambda} = \frac{7 - \cos(\pi/n)}{4[1 + 2\cos(\pi/2n)]^2}, \bar{\zeta} = \frac{[2 + \cos(\pi/2n)]^2}{[1 + 2\cos(\pi/2n)]^2}, \text{ for armchair SWCNTs}(m = n), \tag{11-27}$$

$$\bar{\lambda} = \frac{1}{3 + 3\cos(\pi/n)}, \bar{\zeta} = \frac{[1 + 2\cos(\pi/2n)]^2}{9\cos^2(\pi/2n)}, \text{ for zigzag SWCNTs}(m = 0), \tag{11-28}$$

$$\bar{\lambda} = 1/6, \bar{\zeta} = 1, \text{ for graphene sheets}(m,n \sim \infty). \tag{11-29}$$

The surface Young's modulus and shear modulus are shown in Figure 11-3a. We see that both the moduli increase with increasing tube diameter, approaching a limit value of graphite. For a given tube diameter, the Young's modulus increases but the shear modulus decreases with increasing tube chiral angle. The smaller the tube diameter, the stronger the dependence of the elastic moduli on the tube size and tube chirality. The effects of tube chirality and tube size may be neglected when the tube diameter is larger than 2.0 nm. The present predictions for the initial Young's modulus are in reasonable agreement with those given by various theoretical methods, such as tight bonding calculations [32], lattice dynamics [35], atomistic-based continuum analysis [68], and molecular mechanics approach [60, 94].

The Poisson's ratio against tube diameter is shown in Figure 11-3b. Poisson's ratio decreases with increasing tube diameter, approaching the limit value (0.195) of graphite for large tubes. For a given tube diameter, Poisson's ratio decreases with an increasing in tube chiral angle. Although the present results are in reasonable agreement with some existing data (e.g., [34], there is no unique opinion that is widely accepted for the dependence of Poisson's ratio on the tube diameter. Even completely contrary conclusions were reported in the literatures. Tight binding calculations by Hernandez et al. [32] indicated that Poisson's ratio for armchair tubes increases from 0.247 to 0.256 with increasing tube diameter from (6,6) to (15,15), while for zigzag tubes, Poisson's ratio decreases from 0.275 to 0.270

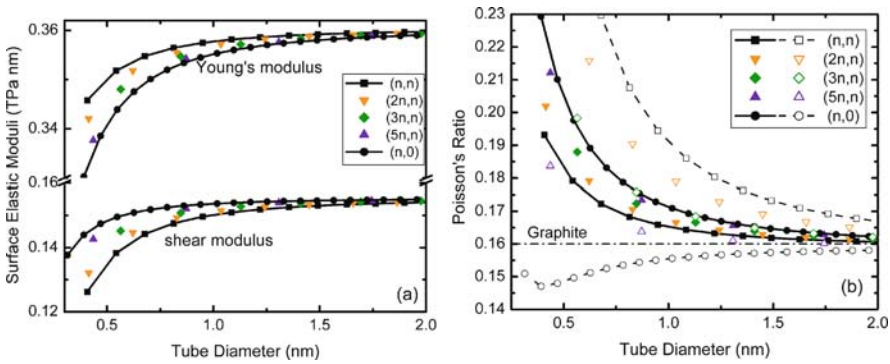


Figure 11-3. Chirality and size dependent elastic properties of SWCNTs: (**a**) Surface Young's modulus and shear modulus; (**b**) Poisson's ratio. The results for the counterfeit Poisson's ratio (which is calculated by $\nu = (Y_S/2–G_S)/G_S$) are also presented for comparison. Apparent discrepancy between the two ratios can be seen. This means that the relationship between Young's modulus and shear modulus in isotropic elastic theory is not retained for SWCNTs

with increasing tube size from (10,0) to (20,0). To the contrary, Popov et al. [35] obtained that, with increasing tube diameter, Poisson's ratio decreases for armchair tubes, but increases for zigzag tubes. The contradiction may be attributed to two aspects. First, there is no sufficient data to eliminate the possible calculating errors during numerical procedures. For instance, only two points are available for zigzag tubes in Hernandez et al.'s work. Second, the relationship between Young's modulus and shear modulus from continuum mechanics is frequently used to extract Poisson's ratio from numerical data, such as in Popov et al.'s work. However, our results indicate that this relationship is NOT retained for a SWCNT. This conclusion agrees with the results from molecular dynamics simulations by Wang et al. [43]. To illustrate this point, we present in 11-3b also the results for Poisson's ratio calculated by the surface Young's modulus and shear modulus via $\nu = (Y_S/2\text{-}G_S)/G_S$ in isotropic elastic theory. We term this Poisson's ratio the *counterfeit* Poisson's ratio. Apparent discrepancy between the two results can be observed. First, for a given tube diameter, the counterfeit Poisson's ratio increases with increasing tube chiral angle, in contradiction with the results directly from the stick-spiral model. Second, with an increase in tube diameter, the counterfeit Poisson's ratio for zigzag tubes (or those near zigzag tubes) with a diameter larger than 0.5 nm increases, while the results directly from the stick-spiral model for tubes with arbitrary chiralities show a decreasing trend.

11.3.3.3. Explicit Expressions for Vibrating Frequencies of Some Raman Modes

Resonance Raman spectroscopy provides a powerful tool to investigate the geometry of SWCNTs. Radial breathing mode (RBM) is one of most important Raman active vibrational modes. Because of its uniqueness, the RBM frequency often

serves as a fingerprint of a free-standing single-walled CNT (SWCNT). Many studies showed that the RBM frequency ω_{RBM} is inversely proportional to the tube diameter d, i.e.

$$\omega_{\mathrm{RBM}} = \frac{A}{d}. \tag{11-30}$$

Different values have been obtained for the constant A (ranging from 218 to 248 cm^{-1} nm, when d is in nm) [99– 105]. Sometimes, an additional term is used to improve the accuracy of the prediction. For instance, a constant B of 10∼30 cm^{-1} is often used to account for the uncertain effects such as those resulted from interaction with the environment. This gives

$$\omega_{\mathrm{RBM}} = \frac{A}{d} + B. \tag{11-31}$$

Although it works well for large tubes, the inverse proportion relationship between the RBM frequency and the tube diameter would be broken down when the tube diameter is smaller than 0.7 nm due to the emergence of the small size and chirality effects in such small tubes [106–111]. Based on the results from numerical simulations, some empirical formulas for the RBM frequency that takes small size and/or chirality effects into account are presented [107, 109, 110, 112, 111]. For example, via a symmetry-adapted nonorthogonal tight-binding model, Popov and Lambin [111] suggested that

$$\omega_{\mathrm{RBM}} = \frac{a_2}{R^{n_2}} + \frac{a_3}{R^{n_3}} \cos\theta, \tag{11-32}$$

where R and θ are respectively the radius and chiral angle of the tube, and a_2, a_3, n_2, and n_3 are constants. Most recently, via a molecular mechanics model, we obtained an analytical relation between the RBM frequency and the structural details of a SWCNT as [17]

$$\omega_{\mathrm{RBM}} = \frac{1}{R}\sqrt{\frac{Y_S' A}{M_0}}, \tag{11-33}$$

with Y_s' the surface circumferential elastic modulus of the tube, A the surface area per atom in graphene plane, and M_0 the mass of a carbon atom. From Eq. (11-33), we can see clearly that the RBM frequency is chirality and size dependent because the surface circumferential elastic modulus Y_s' of a SWCNT is chirality and size dependent while R_0, A_0 and M_0 are all constants. In other words, the chirality and size dependence of the RBM frequency is induced by the chirality and size dependent elastic properties of a SWCNT.

The RBM frequency of a free standing (n, m) SWCNT is presented in Figure 11.4, with an inset showing the product of ω_{RBM} and tube diameter d.

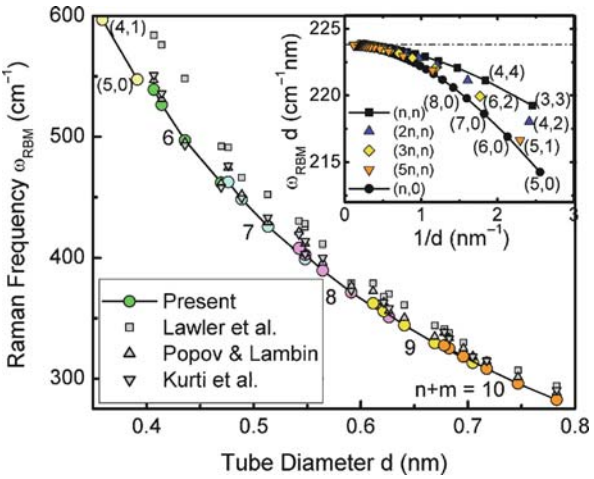


Figure 11-4. Radial breathing mode frequency versus tube diameter for small single-walled carbon nanotubes, with comparisons with first principle, tight binding and ab intio calculations. The inset shows the product of the radial breathing mode frequency and the tube diameter as a function of the inverse tube diameter. It is seen that the inverse proportion law would be broken down for very small tubes

Different families with $n + m =$ const are shown in different colors, starting with an armchair (or close to armchair) tube and ending with a zigzag tube. It is seen that the RBM frequency is dependent not only on the diameter, but also on the chirality. For a given diameter, the tube with a larger chiral angle has a larger RBM frequency. The inversely proportional law may be approximately available for relatively large tubes (i.e., the quantity $\omega_{\mathrm{RBM}}\, d$ tends to a constant with increasing tube diameter). The results from first-principle calculations by Lawler et al. [110], tight binding calculations by Popov and Lambin [111], and ab initio calculations by Kurti et al. [106] are also presented for comparison. Good agreement can be found.

We obtained also the explicit expressions for some G-band Raman vibration modes as [113]

$$\omega_i = \sqrt{\frac{(3 + \mu\kappa)K_r}{M_0}}, \tag{11-34}$$

where κ is a function of tube indexes n and m.

11.3.4. Nonlinear Stick-Spiral Model and its Applications

11.3.4.1. Nonlinear Stick-Spiral Model

The linear stick-spiral model is only suitable for analysis of mechanical behavior of CNTs under small deformations. When a SWCNT subjected to a relatively large strain is considered, nonlinear potentials must be introduced to describe the behavior of atoms far away from their equilibrium positions. In what follows, we demonstrate

how a Morse type potential presented by Belytschko et al. [36] is incorporated into the stick-spiral model. We note that Xiao et al. [94]; (2006) firstly extended Chang and Gao's [5] model to investigate nonlinear behavior of achiral SWCNTs by incorporating this potential. A nonlinear stick-spiral model incorporated the reactive empirical bond-order potential given by Brenner et al. [96] was also developed recently by Duan et al. [95].

The Morse type potential for CNTs can be given by [36]

$$E_{\mathrm{t}} = U_r + U_\theta = \sum_i D_e\{[1 - e^{-\beta(\Delta r_i)}]^2 - 1\} + \frac{1}{2}\sum_j k_\theta(\Delta\theta_j)^2[1 + k_{\mathrm{sextic}}(\Delta\theta_j)^4], \tag{11-35}$$

with $D_e = 0.6031$ nN · nm, $\beta = 26.25$ nm^{-1}, $k_\theta = 1.42$ nN · nm/rad^2, $k_{\mathrm{sextic}} = 0.754$ rad^{-4}. The stretching force resulting from bond elongation and the twisting moment resulting from bond angle variation can be calculated by differentiating the first and the second terms of Eq. (11-35) with respect to bond elongation Δr_i and bond angle variation $\Delta\theta_j$, respectively,

$$F^*(\Delta r_i) = 2\beta D_e(1 - e^{-\beta\Delta r_i})e^{-\beta\Delta r_i}, \tag{11-36}$$

$$M^*(\Delta\theta_j) = k_\theta\Delta\theta_j[1 + 3k_{\mathrm{sextic}}(\Delta\theta_j)^4]. \tag{11-37}$$

11.3.4.2. Mechanical Behavior of SWCNTs Under Large Strains

Mechanical behavior of a SWCNT can be predicted using the nonlinear stick-spiral model. Here we present some of examples for demonstration and detailed procedure can be seen in our previous work [15].

The nonlinear stress-strain relationships are shown in Figure 11-5 for axial loaded SWCNTs with different chiralities but approximately the same diameter. We note that, for the convenience of comparing with the results of others, here the tube wall thickness is simply taken as 0.34 nm, as is most commonly used in the literatures. It is seen that the linear response of a SWCNT is confined within a very small strain region ($\sim\pm3\%$). With increasing tensile strain, the stress-strain curve tends to flatten out, indicating that the SWCNT under tension is strain softening. In contrast to the tensile case, the SWCNT under compression is strain hardening. The tensile stress-strain relations from the present analytical approach agree well with those from the numerical simulations based on molecular dynamics [40, 54, 55] and molecular mechanics [36]. We focus here only on the bond breaking failure mode and the calculations under axial compression are terminated at a relatively small strain (that is assumed not beyond the buckling threshold). The procedure of investigating the CNT buckling using the stick-spiral model can be seen in our previous works [91, 92].

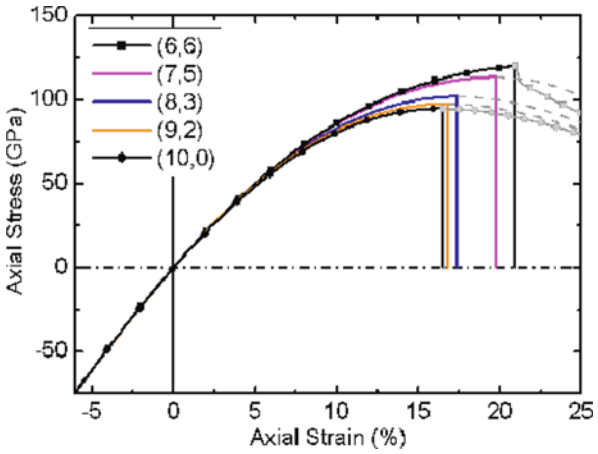


Figure 11-5. Chirality dependent axial stress-strain relationships for SWCNTs with approximately the same diameter

Many efforts have been made to investigate the failure of SWCNTs under axial tension, and two primary failure processes were observed in the studies [49–54, 45, 55]. The first one is brittle fracture (via direct bond breaking), while the second is plastic deformation (via Stone-Wales bond rotation). Brittle failure always leads to a sudden fracture of SWCNTs, whereas plastic deformation may result in necking phenomenon through dislocation evolution [51, 54]. Brittle-to-ductile transition has been extensively discussed by Nardelli et al. [50, 51] and Dumitrica et al. [45] from an atomic point of view. They found that the failure mode of a SWCNT depends not only on tube chirality, but also on applied strain rate and ambient temperature [50–53, 45]. It is obvious that the present simple analytical model could not capture both the two modes because bond reconstruction behavior is not taken into account. Thus it predicts only bond breaking (brittle) failure. That is, once the tube failed (corresponding to the inflection point of a local broken bond), its loading capacity losses immediately [114]. This is the favorable failure mode at low temperatures. On the other hand, the present model is based on an empirical potential [36] that may not be very accurate under very large strains compared to quantum mechanical calculations, as discussed extensively by Zhao et al. [53]. Hence, the present model gives only reference values for tensile strength and failure strain of SWCNTs under brittle fracture, and more accurate results should be obtained by experiments [26, 27], moledular dynamics or *ab initio* calculations [49–52, 36, 53, 38, 40, 56, 54, 45, 55].

The ideal tensile strength of the SWCNT under brittle fracture is the maximum value of the tensile stress which is approached at the inflection point of the stress-strain curve [36, 38, 94, 114]. The failure strain is the applied strain corresponding to the tensile strength. Our results show that the influences of the tube chirality on the tensile strength and failure strain are significant, while the influence of the tube

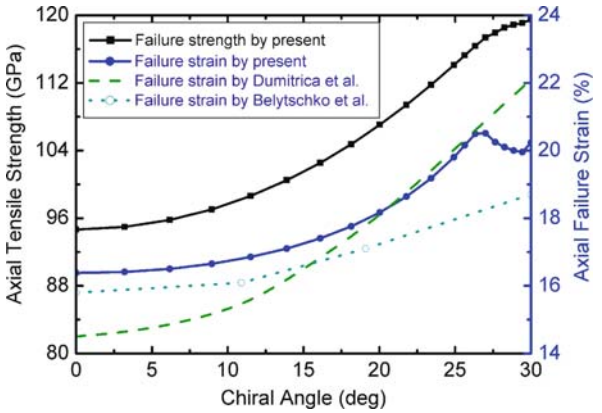


Figure 11-6. Variation of axial tensile strength and failure strain against tube chiral angle. Results from molecular mechanics simulations and quantum mechanics based calculations are also presented for comparison

diameter is ignorable. Both the tensile strength and failure strain approach their limit values when the tube diameter larger than 1 nm.

The axial tensile strength and failure strain of SWCNTs (with diameters larger than 2 nm) versus the tube chirality is shown in Figure 11-6. It is seen that, with increasing tube chiral angle, the tensile strength of a SWCNT increases monotonously. However, the variation of the axial failure strain is quite complicated. With an increase in the chiral angle, the failure strain increases firstly to its maximum value at a chiral angle of about 26°, then decreases to a local minimum at about 29°, and then increases to the value for armchair tubes. The full curve for the failure strain is just like a long dipper. To reveal the physics behind the curve, further study is needed. Results from molecular mechanics simulations by Belytschko et al. [36] and quantum mechanics based calculations by Dumitrica et al. [45] are also presented for comparison and reasonable agreements are found. The tensile strength and the failure strain for armchair tubes (120 GPa and 21%) are about 25 and 30% higher than those for zigzag tubes (95 GPa and 16.4%), respectively. Our predictions are in reasonable agreement with some existing results. Molecular mechanics simulations by Belytschko et al. [36] predicted that the tensile strengths and failure strains of armchair and zigzag tubes are 112 GPa and 18.7%, and 93.5 GPa and 16%, respectively. They found also that the tube size has no effect on the tensile strength. Ogata and Shibutani [38] using a tight binding method gave the tensile strength of about 108 GPa and 114 GPa for zigzag and armchair SWCNTs, respectively. Molecular dynamics simulations by Xiao et al. [40] indicated that the tensile strengths for both armchair and zigzag SWCNTs are about 80 GPa, while the failure strain for zigzag and armchair tubes are 15 and 17%, respectively. Liew et al. [54] obtained by molecular dynamics a tensile strength of 114 GPa and a failure strain of 28% for (10, 10) tubes. We note that the present predicted ideal tensile strength and the failure strain, as well as the mentioned theoretical results, is

significantly higher than some experimental values. For example, Yu et al. [26] measured the tensile strength ranging from 11 to 63 GPa, and the failure strain from 10 to 13% using a scanning electron microscope; Walters et al. [115] using atomic force microscopy obtained the tensile strength of 45 ± 7 GPa for SWCNTs. The lower experimental values might be attributed to the presence of defects, the measuring errors, and so on.

Mechanical behavior of SWCNTs upon radial pressure and torsion, as well as the coupling between different deformations, may also be predicted by the present model. More details can be found in [15]. Here we emphasize in particular that the torsional behavior of a chiral SWCNT is loading direction dependent. The ideal shear strength and failure strain on the tube chirality are shown in Figure 11-7 for SWCNTs with diameters larger than 2 nm. It is seen that, in twisting direction, with increasing chiral angle from zero to $\pi/6$, the ideal shear strength and failure strain increase from the values (115 GPa and 30.1%) for zigzag tubes to their maximum values (146 GPa and 41.3%) at a chiral angle of about $\pi/18$ (9.6 degree in our calculations, slightly dependent on tube diameters), and then decreases to the values (99 GPa and 28.5%) for armchair tubes. In untwisting direction, however, with increasing chiral angle from zero to $\pi/6$, the ideal shear strength and failure strain decrease from the values (115 GPa and 30.1%) for zigzag tubes to their minimum values (95 GPa and 27.8%) at a chiral angle of about $\pi/9$ (18.1° in our calculations, slightly dependent on tube diameters), and then increases to the values (99 GPa and 28.5%) for armchair tubes. The two curves for the shear strength along both the twisting and untwisting directions form a sail-like pattern, so do the curves for the shear failure strain. The ideal shear strength and failure strain for achiral tubes are independent of the loading directions due to their geometrical symmetry, and both the mechanical properties for zigzag tubes are slightly higher than those for armchair tubes. The loading direction dependent buckling behavior of a SWCNT upon torsion is recently investigated by [48] using molecular dynamics simulations.

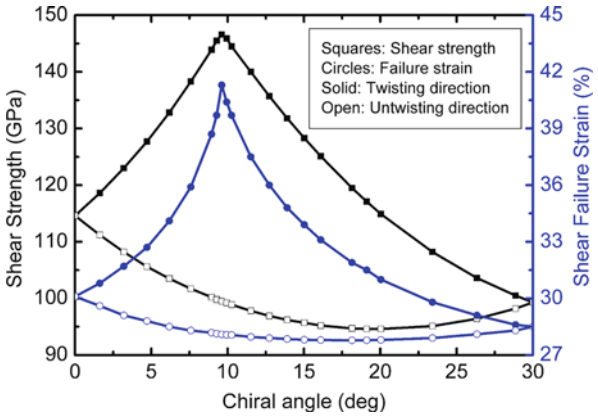


Figure 11-7. Non-monotonous variations of ideal shear strength and failure strain versus tube chiralty. Note that the shear strength and failure strain of a chiral SWCNT in the twisting and untwisting directions are quite different

11.4. CONCLUDING REMARKS

Molecular mechanics methods are based on partition of total energy into contributions associated with bond stretching, angle variation, torsion, inversion, van der Waals, electrostatic, etc. Under most circumstances, molecular mechanics methods can be implemented in atomistic simulations to provide a computational description of various physical phenomena. Under special circumstances, as is shown in this paper, we could simplify a molecular mechanics model sufficiently to obtain analytical solution to certain problems. The analytical solutions nicely complement quantum/atomistic simulations and reveal important parameters in the problem. This might be analogous to parallel development of analytical (e.g., linear elasticity) and numerical methods (e.g., finite element method) in continuum mechanics. Analytical methods and models provide a clear understanding of the physical phenomenon involved and serve as benchmark solutions for numerical methods. On the other hand, numerical methods extend the range of capability of a theory far beyond that reachable by analytical methods alone.

Based on a molecular mechanics concept, a linear stick-spiral model using harmonic potentials and a non-linear stick-spiral model using a Morse type potential are presented for studying mechanical behavior of carbon nanotubes. Closed-form expressions for elastic properties of CNTs can be derived from the linear stick-spiral model. Material properties at different length scales are directly connected via these expressions. Mechanical behavior of CNTs at large strains can be predicted by the nonlinear stick-spiral model. Results from stick-spiral model are in reasonable agreement with the existing numerical results (as well as experimental data), but one needs much less computer time to yield them.

Only a representative atom is used to describe the mechanical behavior of CNTs in the stick-spiral model, which makes it a computationally efficient model. However, this brings also some limitations to the model. For instance, the stochastic effect can not be reflected by this simple model and thus the influence of temperature on the mechanical behavior of CNTs can not be investigated. More robust approaches to account for these effects are importantly necessary.

ACKNOWLEDGMENTS

Financial supports from the National Natural Science Foundation of China (10402019, 10732040), Shanghai Rising-Star Program (05QMX1421), and Shanghai Leading Academic Discipline Project (Y0103) are gratefully acknowledged. The author thanks also all contributors who assisted with developing the stick-spiral model.

APPENDIX

Now let us firstly have an analysis on the triangle ΔAOB in the Figure A1.

In ΔAOB, we have

$$\cos\theta_3 = \frac{1}{2r_1 r_2}\left(r_1^2 + r_2^2 - AB^2\right). \tag{A1}$$

On the other hand,

$$\begin{aligned} AB^2 &= AD^2 + DB^2 = (AE^2 + ED^2 + 2AE \cdot ED\cos\phi_3) + DB^2 \\ &= (r_1\sin\phi_1)^2 + (-r_2\sin\phi_2)^2 - 2r_1r_2\sin\phi_1\sin\phi_2\cos\phi_3 \\ &\quad + (r_1\cos\phi_1 - r_2\cos\phi_2)^2 \\ &= r_1^2 + r_2^2 - 2r_1r_2(\sin\phi_1\sin\phi_2\cos\phi_3 + \cos\phi_1\cos\phi_2). \end{aligned} \tag{A2}$$

Substituting Eq. (A2) into (A1), we obtain

$$\cos\theta_3 = \sin\phi_1\sin\phi_2\cos\varphi_3 + \cos\phi_1\cos\phi_2. \tag{A3}$$

This essentially gives Eq. (11-15).

Next we will show how we get Eq. (11-14).

The angle ω_{13}, which is the torsion angle between the plane (AOE) though r_1 parallel to the nanotube axis and the plane (AOB) of θ_3, can be expressed by the angle BFG. In ΔBFG, we have

$$\cos\omega_{13} = \frac{1}{2BF\cdot FG}\left(BF^2 + FG^2 - BG^2\right). \tag{A4}$$

Note that

$$BF^2 = r_2\sin\theta_3, \tag{A5}$$

$$\begin{aligned} FG^2 &= OF\tan\phi_1 = -r_2\cos\theta_3\tan\phi_1, \\ BG^2 &= OG^2 + OB^2 + 2OG\cdot OB\cos\phi_2 \end{aligned} \tag{A6}$$

$$= \left(\frac{-r_2\cos\theta_3}{\cos\phi_1}\right)^2 + r_2^2 - 2\frac{r_2^2\cos\theta_3\cos\phi_2}{\cos\phi_1}. \tag{A7}$$

Substituting Eqs. (A5), (A6), and (A7) into (A4), we have

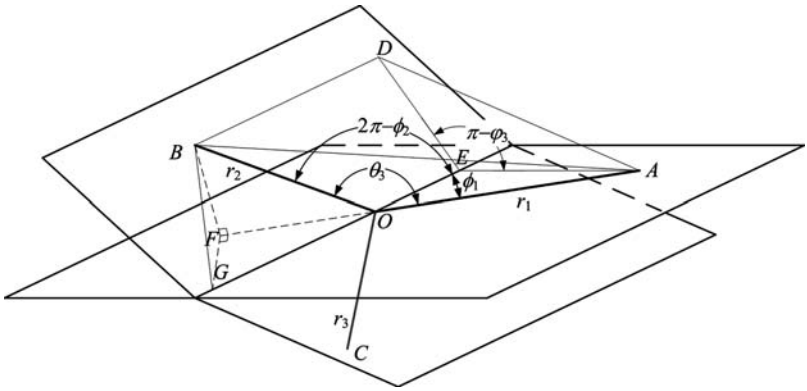


Figure A1. Illustration of some geometrical parameters associated with a carbon atom

$$\cos\omega_{13} = \frac{\cos\theta_3\cos\phi_1 - \cos\phi_2}{\sin\theta_3\sin\phi_1}. \tag{A8}$$

With use of Eq. (A3), Eq. (A8) can be written as

$$\cos\omega_{13} = \frac{\cos\phi_1\sin\phi_2\cos\varphi_3 - \sin\phi_1\cos\phi_2}{\sin\theta_3}. \tag{A9}$$

This actually gives Eq. (11-14).

REFERENCES

1. Shenoy, V.B., Miller, R., Tadmor, E.B., Phillips, R., Ortiz, M., 1998. Quasicontinuum models of interfacial structure and deformation. Phys. Rev. Lett. 80, 742–745.
2. Abraham, F.F., Broughton, J.Q., Bernstein, N., Kaxiras, E., 1998. Spanning the continuum to quantum length scales in a dynamic simulation of brittle fracture. Europhys. Lett. 44, 783–787.
3. Friesecke, G., James, R.D., 2000. A scheme for the passage from atomic to continuum theory for thin films, nanotubes and nanorods. J. Mech. Phys. Solids 48, 1519–1540.
4. Knap, J., Ortiz, M., 2001. An analysis of the quasicontinuum method. J. Mech. Phys. Solids 49, 1899–1923.
5. Chang, T., Gao, H., 2003. Size dependent elastic properties of a single-walled carbon nanotube via a molecular mechanics model. J. Mech. Phys. Solids 51, 1059–1074.
6. Liu, W.K., Karpov, E.G., Zhang, S., Park, H.S., 2004. An introduction to computational nanomechanics and materials. Comput. Methods Appl. Mech. Eng. 193, 1529–1578.
7. Kadowaki, H., Liu, W.K., 2004. Bridging multi-scale method for localization problems. Comput. Methods Appl. Mech. Eng. 193, 3267–3302.
8. Gates, T.S., Odegard, G.M., Frankland, S.J.V., Clancy, T.C., 2005. Computational materials: Multi-scale modeling and simulation of nanostructured materials. Compos. Sci. Technol. 65, 2416–2434.
9. Buryachenko, V.A., Roy, A., Lafdi, K., Anderson, K.L., Chellapilla, S., 2005. Multi-scale mechanics of nanocomposites including interface: Experimental and numerical investigation. Compos. Sci. Technol. 65, 2435–2465.
10. Buehler, M.J., 2006. Large-scale hierarchical molecular modeling of nanostructured biological materials. J. Comput. Theor. Nanosci. 3, 603–623.
11. Danielsson, M., Parks, D.M., Boyce, M.C., 2007. Micromechanics, macromechanics and constitutive modeling of the elasto-viscoplastic deformation of rubber-toughened glassy polymers. J. Mech. Phys. Solids 55, 533–561.
12. Smalley, R.E., Yakobson, B.I., 1998. The future of the fullerenes. Solids State Comm. 107, 597–606.
13. Baughman, R.H., Zakhidov, A.A., de Heer, W.A., 2002. Carbon nanotubes – the route toward applications. Science 297, 787–792.
14. Chang, T., Geng, J., Guo, X., 2005. Chirality- and size- dependent elastic properties of single-walled carbon nanotubes. Appl. Phys. Lett. 87, 251929.
15. Geng, J., Chang, T., 2006. Nonlinear stick-spiral model for predicting mechanical behavior of single-walled carbon nanotubes. Phys. Rev. B 74, 245428.
16. Chang, T., Geng, J., Guo, X., 2006. Prediction of chirality- and size- dependent elastic properties of single-walled carbon nanotubes via a molecular mechanics model. Proc. R. Soc. A 462, 2523–2540.

17. Chang, T., 2007. Explicit solution of the radial breathing mode frequency of single-walled carbon nanotubes. Acta Mech. Sinica 23, 159–162.
18. Iijima, S., 1991. Helical microtubules of graphitic carbon. Nature 354, 56–58.
19. Iijima, S., Ichihashi, T., 1993. Single-shell carbon nanotubes of 1-nm diameter. Nature 363, 603–605.
20. White, C.T., Robertson, D.H., Mintmire, J.W., 1993. Helical and rotational symmetries of nanoscale graphatic tubules. Phys. Rev. B 47, 5485–5488.
21. Treacy, M.M.J., Ebbesen, T.W., Gibson, J.M., 1996. Exceptionally high Young's modulus observed for individual carbon nanotubes. Nature 381, 680.
22. Krishnan, A., Dujardin, E., Ebbesen, T.W., Yianilos, P.N., Treacy, M.M.J., 1998. Young's modulus of single-walled carbon nanotubes. Phys. Rev. B 58, 14013–14019.
23. Poncharal, P., Wang, Z.L., Ugarte, D., de Heer, W.A., 1999. Electrostatic deflections and electromechanical resonances of carbon nanotubes. Science 283, 1513–1516.
24. Wong, E.W., Sheehan, P.E., Lieber, C.M., 1997. Nanobeam mechanics: Elasticity, strength, and toughness of nanorods and nanotubes. Science 277, 1971.
25. Salvetat, J.P., Kulik, A.J., Bonard, J.M., Briggs, G.A.D., Stockli, T., Metenier, K., Bonnamy, S., Beguin, F., Burnham, N.A., Forro, L., 1999. Elastic modulus of ordered and disordered multiwalled carbon nanotubes. Adv. Mater. 11, 161–165.
26. Yu, M.F., Lourie, O., Dyer, M.J., Moloni, K., Kelly, T.F., Ruoff, R.S., 2000. Strength and breaking mechanism of multiwalled carbon nanotubes under tensile load. Science 287, 63.
27. Demczyk, B.G., Wang, Y.M., Cumings, J., Hetman, M., Han, W., Zettl, A., Ritchie, R.O., 2002. Direct mechanical measurement of the tensile strength and elastic modulus of multiwalled carbon nanotubes. Mater. Sci. Eng. A 334, 173–178.
28. Robertson, D.H., Brenner, D.W., Mintmire, J.W., 1992. Energetics of nanoscale graphitic tubules. Phys. Rev. B 45, 12592–12595.
29. Iijima, S., Brabec, C., Maiti, A., Bernholc, J., 1996. Structural flexibility of carbon nanotubes. J. Chem. Phys. 104, 2089–2092.
30. Yakobson, B.I., Brabec, C.J., Bernholc, J., 1996. Nanomechanics of carbon tubes: Instability beyond linear response. Phys. Rev. Lett. 76, 2511–2514.
31. Garg, A., Han, J., Sinnott, S.B., 1998. Interactions of carbon nanotuble proximal probe tips with diamond and graphene. Phys. Rev. Lett. 81, 2260–2263.
32. Hernandez, E., Goze, C., Bernier, P., Rubio, A., 1998. Elastic properties of C and BxCyNz composite nanotubes. Phys. Rev. Lett. 80, 4502–4505.
33. Yao, N., Lordi, V., 1998. Young's modulus of single-walled carbon nanotubes. J. Appl. Phys. 84, 1939–1943.
34. Sanchez-Portal, D., Artacho, E., Soler, J.M., Rubio, A., Ordejon, P., 1999. Ab initio structural, elastic, and vibrational properties of carbon nanotubes. Phys. Rev. B 59, 12678–12688.
35. Popov, V.N., Van Doren, V.E., Balkanski, M., 2000. Elastic properties of single-walled carbon nanotubes. Phys. Rev. B 61, 3078–3084.
36. Belytschko, T., Xiao, S.P., Schatz, G.C., Ruoff, R.S., 2002. Atomistic simulations of nanotube fracture. Phys. Rev. B 65, 235430.
37. Krenn, R.C., Roundy, D., Cohen, M.L., Chrzan, D.C., Morris, J.J.W., 2002. Connecting atomistic and experimental estimates of ideal strength. Phys. Rev. B 65, 134111.
38. Ogata, S., Shibutani, Y., 2003. Ideal strength and band gap of single-walled carbon nanotubes. Phys. Rev. B 68, 165409.
39. Gartstein, Y.N., Zakhidov, A.A., Baughman, R.H., 2003. Mechanical and electromechanical coupling in carbon nanotube distortions. Phys. Rev. B 68, 115415.
40. Xiao, T., Xu, X., Liao, K., 2004. Characterization of nonlinear elasticity and elastic instability in single-walled carbon nanotubes. J. Appl. Phys. 95, 8145–8148.

41. Liew, K.M., Wong, C.H., He, X.Q., Tan, M.J., Meguid, S.A., 2004. Nanomechanics of single and multi walled carbon nanotubes. Phys. Rev. B 69, 115429.
42. Wang, Y., Wang, X.X., Ni, X.G., 2004. Atomistic simulation of the torsion deformation of carbon nanotubes. Model. Simul. Mater. Sci. Eng. 12, 1099–1107.
43. Wang, L.F., Zheng, Q.S., Liu, J.Z., Jiang, Q., 2005. Size dependence of the thin-shell model for carbon nanotubes. Phys. Rev. Lett. 95, 105501.
44. Liang, H.Y., Upmanyu, M., 2006. Axial-strain-induced torsion in single-walled carbon nanotubes. Phys. Rev. Lett. 96, 165501.
45. Dumitrica, T., Hua, M., Yakobson, B.I., 2006. Symmetry-, time-, and temperature-dependent strength of carbon nanotubes. Proc. Nat. Ac. Sci. USA 103, 6105–6109.
46. Chang, T., Hou, J., 2006. Molecular dynamics simulations on buckling of multiwalled carbon nanotubes under bending. J. Appl. Phys. 100, 114327.
47. Chang, T., Hou, J., Guo, X., 2006. Reversible mechanical bistability of single-walled carbon nanotubes under axial strain. Appl. Phys. Lett. 88, 211906.
48. Chang, T., 2007. Torsional behavior of chiral single-walled carbon nanotubes is loading direction dependent. Appl. Phys. Lett. 90, 201910.
49. Yakobson, B.I., Campbell, M.P., Brabec, C.J., Bernholc, J., 1997. High strain rate fracture and C-chain unraveling in carbon nanotubes. Comput. Mater. Sci. 8, 341–348.
50. Nardelli, M.B., Yakobson, B.I., Bernholc, J., 1998. Mechanism of strain release in carbon nanotubes. Phys. Rev. B 57, R4277.
51. Nardelli, M.B., Yakobson, B.I., Bernholc, J., 1998. Brittle and ductile behavior in carbon nanotubes. Phys. Rev. Lett. 81, 4656.
52. Nardelli, M.B., Fattebert, J.L., Orlikowski, D., Roland, C., Zhao, Q., Bernholc, J., 2000. Mechanical properties, defects and electronic behavior of carbon nanotubes. Carbon 38, 1703–1711.
53. Zhao, Q., Nardelli, M.B., Bernholc, J., 2002. Ultimate strength of carbon nanotubes: A theoretical study. Phys. Rev. B 65, 144105.
54. Liew, K.M., He, X.Q., Wong, C.H., 2004. On the study of elastic and plastic properties of multi-walled carbon nanotubes under axial tension using molecular dynamics simulation. Acta Materialia 52, 2521–2527.
55. Liew, K.M., Wong, C.H., Tan, M.J., 2006. Tensile and compressive properties of carbon nanotube bundles. Acta Materialia 54, 225–231.
56. Shibutani, Y., Ogata, S., 2004. Mechanical itegrity of single-walled carbon nanotubes for bending and torsion. Mod. Sim. Mater. Sci. Eng. 12, 599–610.
57. Zhang, Y.Y., Tan, V.B.C., Wang, C.M., 2006. Effect of chirality on buckling behavior of single-walled carbon nanotubes. J. Appl. Phys. 100, 074304–074306.
58. Duan, X., Tang, C., Zhang, J., Guo, W., Liu, Z., 2007. Two distinct buckling modes in carbon nanotube bending. Nano Lett. 7, 143–148.
59. Arroyo, M., Belytschko, T., 2002. An atomistic-based finite deformation membrane for single layer crystalline films. J. Mech. Phys. Solids 50, 1941–1977.
60. Li, C.Y., Chou, T.W., 2003. A structural mechanics approach for the analysis of carbon nanotubes. Int. J. Solids Struct. 40, 2489–2499.
61. Wagner, G.J., Liu, W.K., 2003. Coupling of atomistic and continuum simulations using a bridging scale decomposition. J. Comput. Phys. 190, 249–274.
62. Pantano, A., Parks, D.M., Boyce, M.C., 2004. Mechanics of deformation of single- and multi-wall carbon nanotubes. J. Mech. Phys. Solids 52, 789–821.
63. Qian, D., Wagner, G.J., Liu, W.K., 2004. A multiscale projection method for the analysis of carbon nanotubes. Comput. Methods Appl. Mech. Eng 193, 1603–1632.

64. Zhang, H.W., Wang, J.B., Guo, X., 2005. Predicting the elastic properties of single-walled carbon nanotubes. J. Mech. Phys. Solids 53, 1929–1950.
65. Behdinan, K., Xu, Y.G., Fawaz, Z., 2005. Molecular element method (MEM) for multi-scale modeling and simulations of nano/micro-systems. Trans. CSME 29, 403–421.
66. Liu, B., Jiang, H., Huang, Y., Qu, S., Yu, M.F., Hwang, K.C., 2005. Atomic-scale finite element method in multiscale computation with applications to carbon nanotubes. Phys. Rev. B 72, 035435.1–035435.8.
67. Wang, M., Zhang, X., Lu, M.W., 2005. Nonlinear membrane-spring model for carbon nanotubes. Phys. Rev. B 72, 205403–205407.
68. Guo, X., Wang, J.B., Zhang, H.W., 2006. Mechanical properties of single-walled carbon nanotubes based on higher order Cauchy-Born rule. Int. J. Solids Struct. 43, 1276–1290.
69. Cao, G., Chen, X., 2006. Buckling of single-walled carbon nanotubes upon bending: Molecular dynamics simulations and finite element method. Phys. Rev. B 73, 155435.
70. Dumitrica, T., James, R.D., 2007. Objective molecular dynamics. J. Mech. Phys. Solids 55, 2206–2236.
71. Wang, M., Zhang, X., Zheng, Q., Liu, Y., 2007. A membrane-spring model for carbon nanotubes with van der Waals interaction between non-bonded atoms. Nanotechnology 18, 375706.
72. Wu, J., Hwang, K.C., Huang, Y., 2008. An atomistic-based finite-deformation shell theory for single-wall carbon nanotubes. J. Mech. Phys. Solids 56, 279–292.
73. Ru, C.Q., 2000. Column buckling of multiwalled carbon nanotubes with interlayer radial displacements. Phys. Rev. B 62, 16962.
74. Wang, C.Y., Ru, C.Q., Mioduchowski A., 2003. Axially compressed buckling of pressured multiwall carbon nanotubes. Int. J. Solids Struct. 40, 3893–3911.
75. Han, Q., Lu, G.X., 2003. Torsional buckling of a double-walled carbon nanotube embedded in a elastic medium. Eur. J. Mech. A Solids 22, 875–883.
76. Shen, H.S., 2004. Postbuckling prediction of double-walled carbon nanotubes under hydrostatic pressure. Int. J. Solids Struct. 41, 2643–2657.
77. Kitipornchai, S., He, X.Q., Liew, K.M., 2005. Continuum model for the vibration of multilayered graphene sheets. Phys. Rev. B 72, 075443.
78. He, X.Q., Kitipornchai, S., Liew, K.M., 2005. Buckling analysis of multi-walled carbon nanotubes: A continuum model accounting for van der Waals interaction.. J. Mech. Phys. Solids 53, 303–326.
79. Wang, X., Yang, H.K., 2006. Bending stability of multiwalled carbon nanotubes. Phys. Rev. B 73, 085409.
80. Longhurst, M.J., Quirke, N., 2006. The environmental effect on the radial breathing mode of carbon nanotubes in water. J. Chem. Phys. 124, 234708.
81. Shen, H.S., Zhang, C.L., 2006. Postbuckling prediction of axially loaded double-walled carbon nanotubes with temperature dependent properties and initial defects. Phys. Rev. B 74, 035410.
82. Wang, Q., Varadan, V.K., Quek, S.T., 2006. Small scale effect on elastic buckling of carbon nanotubes with nonlocal continuum models. Phys. Lett. A 357, 130–135.
83. Zhang, Y.Q., Liu, G.R., Han, X., 2006. Effect of small length scale on elastic buckling of multi-walled carbon nanotubes under radial pressure. Phys. Lett. A 349, 370–376.
84. Tylikowski, A., 2008. Instability of thermally induced vibrations of carbon nanotubes. Arch. Appl. Mech. 78, 49–60.
85. Wang, C.Y., Ru, C.Q., Mioduchowski, A., 2004. Applicability and limitations of simplified eleastic shell equations for carbon nanotubes. J. Appl. Mech ASME 71, 622–631.
86. Ru, C.Q., 2000b. Effective bending stiffness of carbon nanotubes. Phys. Rev. B 62, 9973.
87. Odegard, G.M., Gates, T.S., Nicholson, L.M., Wise, K.E., 2002. Equivalent-continuum modeling of nano-structured materials. Compos. Sci. Technol. 62, 1869–1880.

88. Wang, Q., 2004. Effective in-plane stiffness and bending rigidity of armchair and zigzag carbon nanotubes. Int. J. Solids Struct. 41, 5451–5461.
89. Shen, L., Li, J., 2004. Transversely isotropic elastic properties of single-walled carbon nanotubes. Phys. Rev. B 69, 045414.
90. Leung, A.Y.T., Guo, X., He, X.Q., Kitipornchai, S., 2005. A continuum model for zigzag single-walled carbon nanotubes. Appl. Phys. Lett. 86, 083110.
91. Chang, T., Guo, W., Guo, X., 2005. Buckling of multi-walled carbon nanotubes under axial compression and bending via a molecular mechanics model. Phys. Rev. B 72, 064101.
92. Chang, T., Li, G., Guo, X., 2005. Elastic axial buckling of carbon nanotubes via a molecular mechanics model. Carbon 43, 287–294.
93. Wang, Q., Duan, W.H., Liew, K.M., He, X.Q., 2007. Inelastic buckling of carbon nanotubes. Appl. Phys. Lett. 90, 033110–033113.
94. Xiao, J.R., Gama, B.A., Gillespie, J.W., 2005. An analytical molecular structural mechanics model for the mechanical properties of carbon nanotubes. Int. J. Solids Struct. 42, 3075–3092.
95. Duan, W.H., Wang, Q., Liew, K.M., He, X.Q., 2007. Molecular mechanics modeling of carbon nanotube fracture. Carbon 45, 1769–1776.
96. Brenner, D.W., Shenderova, O.A., Harrison, J.A., Stuart, S.J., Ni, B., Sinnott, S.B., 2002. A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. J. Phys. Condens. Matter. 14, 783–802.
97. Allinger, N.L., 1977. Conformational analysis 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. J. Am. Chem. Soc. 99, 8127–8134.
98. Leach, A.R., 1996. Molecular Modelling: Principles and Applications. Addison Wesley Longman Limited, London.
99. Jishi, R.A., Venkataraman, L., Dresselhaus, M.S., Dresselhaus, G., 1993. Phonon modes in carbon nanotubules. Chem. Phys. Lett. 209, 77.
100. Kurti, J., Kresse, G., Kuzmany, H., 1998. First-principles calculations of the radial breathing mode of single-wall carbon nanotubes. Phys. Rev. B 58, R8869.
101. Bandow, S., Asaka, S., Saito, Y., Rao, A.M., Grigorian, L., Ritchter, E., Eklund, P.C., 1998. Effect of the growth temperature on the diameter distribution and chirality of single-wall carbon nanotubes. Phys. Rev. Lett. 80, 3779–3782.
102. Sanchez-Portal, D., Artacho, E., Soler, J.M., Rubio, A., Ordejon, P., 1999. Ab initio structural, elastic, and vibrational properties of carbon nanotubes. Phys. Rev. B 59, 12678–12688.
103. Jorio, A., Saito, R., Hafner, J.H., Lieber, C.M., Hunter, M., McClure, T., Dresselhaus, G., Dresselhaus, M.S., 2001. Structural (n,m) determination of isolated single wall carbon nanotubes by resonant Raman scattering. Phys. Rev. Lett. 86, 1118.
104. Dobardzic, E., Milosevic, I., Nikolic, B., Vukovic, T., Damnjanovic, M., 2003. Single-wall carbon nanotubes phonon spectra: Symmetry-based calculations. Phys. Rev. B 68, 045408.
105. Longhurst, M.J., Quirke, N., 2005. The radial breathing mode of carbon nanotubes. Mol. Simul. 31, 135.
106. Kurti, J., Zolyomi, V., Kertesz, M., Sun, G., 2003. The geometry and the radial breathing mode of carbon nanotubes: Beyond the ideal behaviour. New J Phys. 5, 125.
107. Telg, H., Maultzsch, J., Reich, S., Hennrich, F., Thomsen, C., 2004. Chirality distribution and transition energies of carbon nanotubes. Phys. Rev. Lett. 93, 177401.
108. Damnjanovic, M., Dobardzic, E., Milosevic, I., 2004. Chirality dependence of the radial breathing mode: A simplemodel. J. Phys. Condens. Matter. 16, L505–L508.
109. Meyer, J.C., Paillet, M., Michel, T., Moreac, A., Neumann, A., Duesberg, G.S., Roth, S., Sauvajol, J., 2005. Raman modes of index-identified freestanding single-walled carbon nanotubes. Phys. Rev. Lett. 95, 217401.

110. Lawler, H.M., Areshkin, D., Mintmire, J.W., White, C.T., 2005. Radial-breathing mode frequencies for single-walled carbon nanotubes of arbitrary chirality: First-principles calculations. Phys. Rev. B 72, 233403.
111. Popov, V.N., Lambin, P., 2006. Radius and chirality dependence of the radial breathing mode and the G-band phonon modes of single-walled carbon nanotubes. Phys. Rev. B 73, 085407.
112. Xiao, Y., Li, Z.M., Yan, X.H., Zhang, Y., Mao, Y.L., Yang, Y.R., 2005. Curvature effect on the radial breathing modes of single-walled carbon nanotubes. Phys. Rev. B 71, 233405.
113. Li, L., Chang, T., 2008. Explicit solutions for G-band Raman frequencies of single-walled carbon nanotubes. Acta Mech. Solida Sinica, (in press).
114. Xiao, J.R., Gillespie, J., 2006. Nonlinear deformation and progressive failure of multiwalled carbon nanotubes under internal radial pressure. Phys. Rev. B 74, 155404–155407.
115. Walters, D.A., Ericson, L.M., Casavant, M.J., 1999. Elastic strain of freely suspended single-wall carbon nanotube ropes. Appl. Phys. Lett 74(25), 3803–3805.

CHAPTER 12

POTENTIALS FOR VAN DER WAALS INTERACTION IN NANO-SCALE COMPUTATION

J. XIAO[1], W. ZHOU[2], Y. HUANG[1,3], J.M. ZUO[4], AND K.C. HWANG[5]

[1]*Department of Mechanical Engineering, Northwestern University, Evanston, IL 60208, USA*
[2]*Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Champaign, IL 61801, USA*
[3]*Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL 60208, USA, e-mail: y-huang@northwestern.edu*
[4]*Department of Materials Science and Engineering and Materials Research Laboratory, University of Illinois at Urbana-Champaign, Champaign, IL 61801, USA*
[5]*Department of Engineering Mechanics, Tsinghua University, Beijing 100084, China*
Correspondence to Y. Huang, y-huang@northwestern.edu

Abstract: Van der Waals interaction is important to the materials behavior at the nanoscale. Two potentials for van der Waals interactions in carbon, namely the Lennard-Jones potential and registry-dependent interlayer potential, are compared through graphenes and carbon nanotubes. The registry-dependent interlayer potential has stronger lattice registry effect than the Lennard-Jones potential, and agrees better with first-principles computation. However, the deformation of carbon nanotubes predicted by Lennard-Jones potential shows better agreement with experiments than that by registry-dependent interlayer potential

Keywords: van der Waals interaction, Potential, Computation

12.1. INTRODUCTION

Van der Waal interaction is much weaker than chemical bonds, but plays a fundamental role in nanotechnology. Chopra et al. [1] first observed fully collapsed carbon nanotubes induced by van der Waals interaction, the mechanism within which was extensively studied by experiments [2–4] and computations. [5–8] The collapse of carbon nanotubes changes the electrical properties, [9] which has many potential applications such as field effect transistors, [10] mechanical nanoswitchs, [11] and nanoscale pressure sensors. [12]

Recently, van der Waals was found responsible for dry adhesion of gecko setae, [13] which inspired the carbon nanotube-based synthetic gecko tapes [14] and biocompatible tissue adhesives [15] with the potential applications in microelctronics, robotics and medical therapies. [14, 15]

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 323–333.
DOI 10.1007/978-1-4020-9785-0_12,

In order to investigate the effect of van der Waals interaction for carbon, an accurate potential describing van der Waals interaction is needed. This chapter compares the widely used the Lennard-Jones potential [16] and recently developed registry-dependent interlay potential [17] for van der Waals interaction.

12.2. POTENTIALS FOR VAN DER WAALS INTERACTION

12.2.1. The Lennard-Jones Potential

The Lennard-Jones potential is widely used to characterize the van der Waals interactions. The interaction energy between two atoms is [16]

$$v(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{12-1}$$

where r is the distance between the two atoms, ε is the well depth and σ is related to the equilibrium distance r_0 by $r_0 = 2^{1/6}\sigma$. For van der Waals interactions among carbon atoms, the parameters are $\varepsilon = 2.39$ meV and $\sigma = 0.342$ nm. [16] The above potential is a two-body interatomic potential, and has shown good agreement with experiments for van der Waals interactions among atoms. [8, 18]

12.2.2. The Registry-Dependent Interlayer Potential

Recently, Kolmogorov et al. [17] pointed out that the Lennar-Jones potential underestimated the lattice registry effect in graphitic systems, and proposed an improved classical interatomic potential (registry-dependent interlayer potential) to characterize van der Waals interactions, which has the expression as [17]

$$\begin{gathered} v\left(\mathbf{r}_{ij},\mathbf{n}_i,\mathbf{n}_j\right) = e^{-\lambda\left(r_{ij}-z_0\right)}\left[C + f\left(\rho_{ij}\right) + f\left(\rho_{ji}\right)\right] - A\left(\frac{r_{ij}}{z_0}\right)^{-6}, \\ \rho_{ij}^2 = r_{ij}^2 - \left(\mathbf{n}_i\mathbf{r}_{ij}\right)^2, \\ \rho_{ji}^2 = r_{ij}^2 - \left(\mathbf{n}_j\mathbf{r}_{ij}\right)^2, \\ f(\rho) = e^{-(\rho/\delta)^2}\sum C_{2n}\left(\rho/\delta\right)^{2n}, \end{gathered} \tag{12-2}$$

where r_{ij} is the distance between atoms i and j, and $\mathbf{n}_k$ is the normal to the sp^2 plane at the position of atom k, the proper definition of which was given by Kolmogorov et al. [17]. The parameters for graphitic systems are $C_0 = 15.71$ meV, $C_2 = 12.29$ meV, $C_4 = 4.933$ meV, $C = 3.030$ meV, $A = 10.238$ meV, $\delta = 0.0578$ nm, $z_0 = 0.334$ nm and $\lambda = 36.29$ nm^{-1} [17]. This registry-dependent interlayer potential shows good agreement with the first-principles computations on lattice registry effect of graphitic systems [17].

12.3. COMPUTATIONAL METHOD

Atomic-scale finite element method (AFEM) [19, 20] is introduced in this section, which is used to study the graphitic systems for the comparison of the Lennard-Jones and registry-dependent interlayer potentials.

The total energy for a system of N atoms could be evaluated by interatomic potentials, and is denoted by $E_{\text{tot}}(\boldsymbol{x})$, where $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)^{\text{T}}$, $\boldsymbol{x}_i$ is the position of atom i. The state of minimal energy corresponds to

$$\frac{\partial E_{\text{tot}}}{\partial \boldsymbol{x}} = 0. \tag{12-3}$$

The Taylor expansion of E_{tot} around an initial guess $\boldsymbol{x}^{(0)} = (\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_2^{(0)}, \ldots, \boldsymbol{x}_N^{(0)})^T$ of the equilibrium state gives

$$E_{\text{tot}}(\boldsymbol{x}) \approx E_{\text{tot}}(\boldsymbol{x}^{(0)}) + \left.\frac{\partial E_{\text{tot}}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(0)}} \cdot (\boldsymbol{x} - \boldsymbol{x}^{(0)}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^{(0)})^T \cdot \left.\frac{\partial^2 E_{\text{tot}}}{\partial \boldsymbol{x} \partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(0)}} \cdot (\boldsymbol{x} - \boldsymbol{x}^{(0)}). \tag{12-4}$$

Its substitution into Eq. (12-3) yields the following governing equation for the displacement $\mathbf{u} = \boldsymbol{x} - \boldsymbol{x}^{(0)}$,

$$\mathbf{K}\mathbf{u} = \mathbf{P}, \tag{12-5}$$

where $\mathbf{K} = \left.\partial^2 E_{\text{tot}} / \partial \boldsymbol{x} \partial \boldsymbol{x}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(0)}}$, and $\mathbf{P} = -\left.\partial E_{\text{tot}} / \partial \boldsymbol{x}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(0)}}$ which is the steepest descent direction of E_{tot}. The above equation is identical to the governing equation in continuum finite element method (FEM) if atoms are replaced by FEM nodes. In fact, $\mathbf{K}$ and $\mathbf{P}$ are called the stiffness matrix and nonequilibrium force vector in FEM, respectively. For a nonlinear system, Eq. (12-5) is solved iteratively until $\mathbf{P}$ reaches zero.

Since each atom interacts only with finite neighbor atoms (but not necessarily nearest-neighbor atoms), the first and second order derivatives, $\partial E_{\text{tot}}/\partial \boldsymbol{x}$ and $\partial^2 E_{\text{tot}}/\partial \boldsymbol{x}\boldsymbol{x}$, of E_{tot} with respect to atom i can be calculated via a small subset of atoms including atom i and its neighbor atoms. Such a subset of atoms is called an element in AFEM, and the composition of element depends on the atomic structure and nature of atomistic interactions, as to be discussed in the example shown in Figure 12-1. The contribution from this element to $\mathbf{K}$ is called the element stiffness matrix $\mathbf{K}_i^{\text{element}}$ such that $\mathbf{K}$ is the assemble of all element stiffness matrices.

A carbon nanotube (CNT) as shown in Figure 12-1 is used as an example to illustrate the AFEM element. Figure 12-1 shows a three-dimensional AFEM element for CNT containing a central atom (No. 1), three nearest neighbor (Nos. 2,5,8) and six second-nearest neighbor atoms (Nos. 3,4,6,7,9, and 10). The atomistic interaction among carbon atoms is characterized by multibody interatomic potentials [21] which indicate that each atom (No. 1) on a CNT interacts with not only three nearest-neighbor atoms but also six second-nearest neighbor atoms via the bond angle change. For example, energy stored in an atomic bond between atoms 1

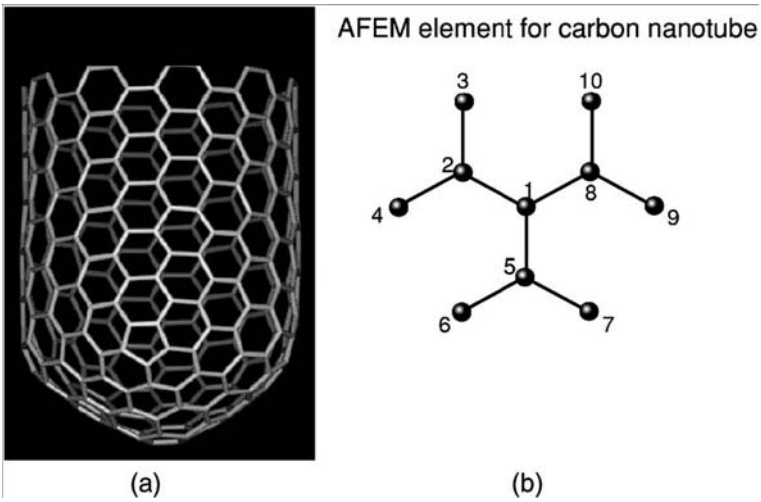


Figure 12-1. (**a**) Schematic diagram of a single wall carbon nanotube; (**b**) an atomic-scale finite element for carbon nanotubes

and 2 depends on not only the bond length but also angles with neighbor bonds 1–5, 1–8, 2–3, and 2–4, reflecting the multibody nature of atomistic interactions. Therefore, the position change of central atom 1 influences energy stored in nine atomic bonds within this element shown in Figure 12-1. Such an element captures interactions between the central atom and all neighbor atoms, and is used to calculate $\partial E_{\text{tot}}/\partial \boldsymbol{x}$ and $\partial^2 E_{\text{tot}}/\partial \boldsymbol{x}$ associated with the central atom. It gives the following element stiffness matrix $\mathbf{K}^{\text{element}}$ and nonequilibrium force vector $\mathbf{P}^{\text{element}}$:

$$\mathbf{K}^{\text{element}} = \begin{bmatrix} \left(\dfrac{\partial^2 E_{\text{tot}}}{\partial \boldsymbol{x}_1 \partial \boldsymbol{x}_1}\right)_{3\times 3} & \left(\dfrac{1}{2}\dfrac{\partial^2 E_{\text{tot}}}{\partial \boldsymbol{x}_1 \partial \boldsymbol{x}_i}\right)_{3\times 27} \\ \left(\dfrac{1}{2}\dfrac{\partial^2 E_{\text{tot}}}{\partial \boldsymbol{x}_i \partial \boldsymbol{x}_1}\right)_{27\times 3} & (0)_{27\times 27} \end{bmatrix}, \tag{12-6}$$

$$\mathbf{P}^{\text{element}} = \begin{bmatrix} \left(-\dfrac{\partial U_{\text{tot}}}{\partial \boldsymbol{x}_1}\right)_{3\times 1} \\ (0)_{27\times 1} \end{bmatrix}, \tag{12-7}$$

Therefore, each row in the stiffness matrix $\mathbf{K}$ assembled from element stiffness matrices has at most 30 nonzero components (since each element has tens atoms) such that $\mathbf{K}$ is sparse and the number of nonzero components is on the order N, i.e., O(N), where N is the total number of atoms in the system. It is important to point out that AFEM does not involve any approximations of continuum FEM (e.g., interpolation), and gives accurate results.

Since the effort to compute **K** and **P** is O(N) and the effort to solve $\mathbf{Ku} = \mathbf{P}$ in Eq. (12-5) is also O(N) due to the sparseness of **K,** AFEM is an order-N computational method, and thus is much faster than the well-known conjugate gradient method. [19, 20]

12.4. COMPARISON BETWEEN THE TWO POTENTIALS

The Lennard-Jones potential [16] and registry-dependent interlayer potential [17] with the parameters given in Section 12.2 are implemented in AFEM to account for van der Waals interaction among atoms within graphitic systems, while the covalent bonds are characterized by the second generation Tersoff-Brenner potential. [21]

12.4.1. On the Lattice Registry Effect

The effect of lattice registry between two parallel graphenes is studied first to compare the Lennard-Jones potential and registry-dependent interlayer potential. Since there are two different registry states, AB stacking and AA stacking as shown in Figure 12-2, between the parallel graphenes, their energies given by the two potentials are compared separately.

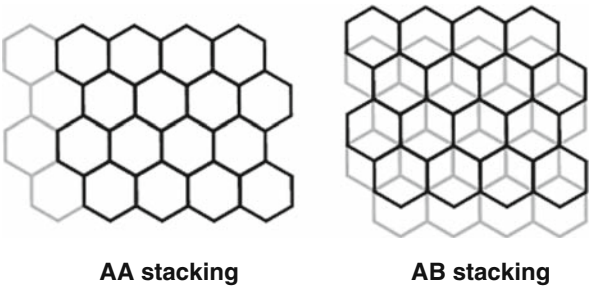


Figure 12-2. Schematic diagram of AA and AB stacking between two graphenes

(i) *AB stacking*. Figure 12-3 shows the van der Waals interlayer energy per carbon atom between two AB-stacking parallel graphenes versus their distance. The curves given by the Lennar-Jones potential and by the registry-dependent interlayer potential are rather close, and both give an equilibrium distance around 0.34 nm.

(ii) *AA stacking*. The van der Waals interlayer energy per carbon atom between two AA-stacking parallel graphenes is shown in Figure 12-4 versus the interlayer distance. As the distance decreases, the van der Waals energy given by the registry-dependent interlayer potential increases much faster than that given the by Lennar-Jones potential. In addition, the equilibrium distances given by the two potentials are also different, 0.344 nm for the Lennard-Jones potential and 0.362 nm for the registry-dependent interlayer potential, respectively.

The lattice registry effect of two adjacent graphite layers is determined by the difference between the van der Waals energies at AA and AB stacking, which is shown

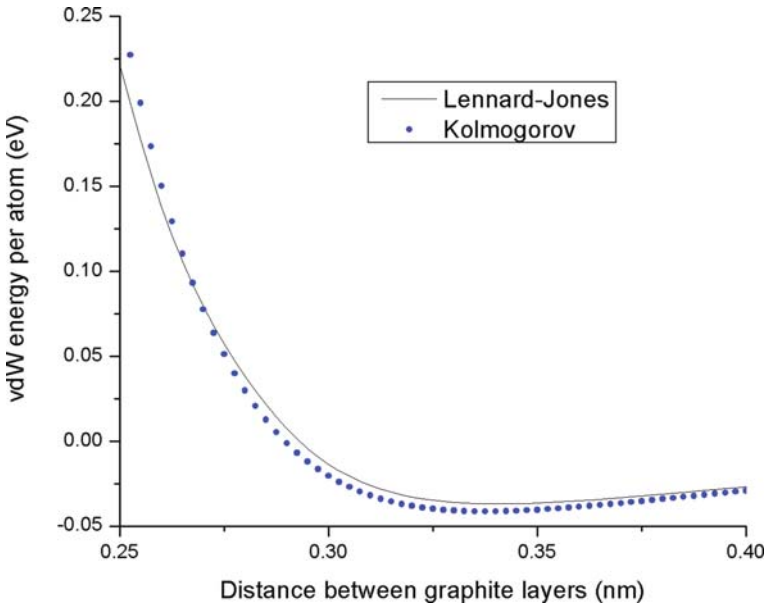


Figure 12-3. The van der Waals interlayer energy per carbon atom between two AB-stacking parallel graphenes versus their distance

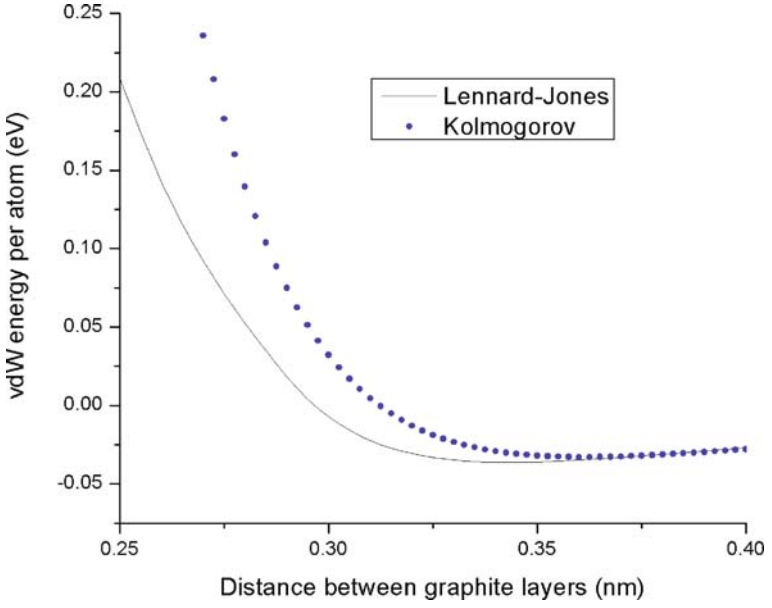


Figure 12-4. The van der Waals interlayer energy per carbon atom between two AA-stacking parallel graphenes versus their distance

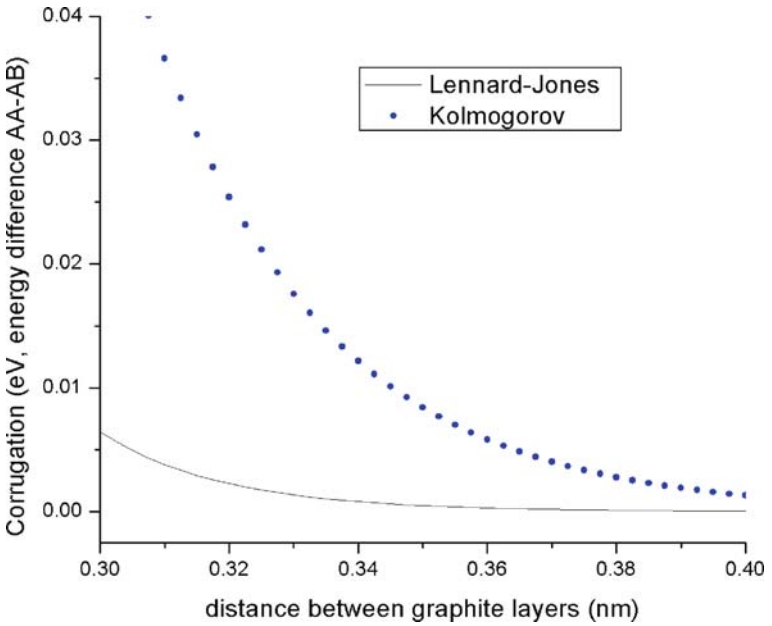


Figure 12-5. The difference in the van der Waals energy between AA and AB stacking given by Lennard-Jones potential and registry-dependent interlayer potential

in Figure 12-5. The registry-dependent interlayer potential shows much stronger lattice registry effect than the Lennar-Jones potential, and it has better agreement with the first-principles calculations. [17] This suggests that the Lennar-Jones potential may underestimate the lattice registry effect in graphitic systems.

12.4.2. On the Deformation of Carbon Nanotubes

The effect of Lennard-Jones potential and registry-dependent interlayer potential on the deformation of carbon nanotubes is compared in this section. The van der Waals energy versus distance between two (10, 10) armchair single-wall carbon nanotubes is shown in Figure 12-6 for the Lennard-Jones potential and registry-dependent interlayer potential. There are 800 atoms in each tube. The equilibrium distance given by the registry-dependent interlayer potential, 0.31 nm, is slightly smaller that that given by the Lennard-Jones potential, 0.32 nm. Furthermore, the registry-dependent interlayer potential gives a larger well depth than the Lennar-Jones potential, and therefore larger deformation of carbon nanotubes.

The effect of Lennar-Jones potential and registry-dependent interlayer potential on the deformation of carbon nanotubes is also investigated via a bundle of two chiral single-wall carbon nanotubes (17,5) and (19,11). Figure 12-7 shows the cross

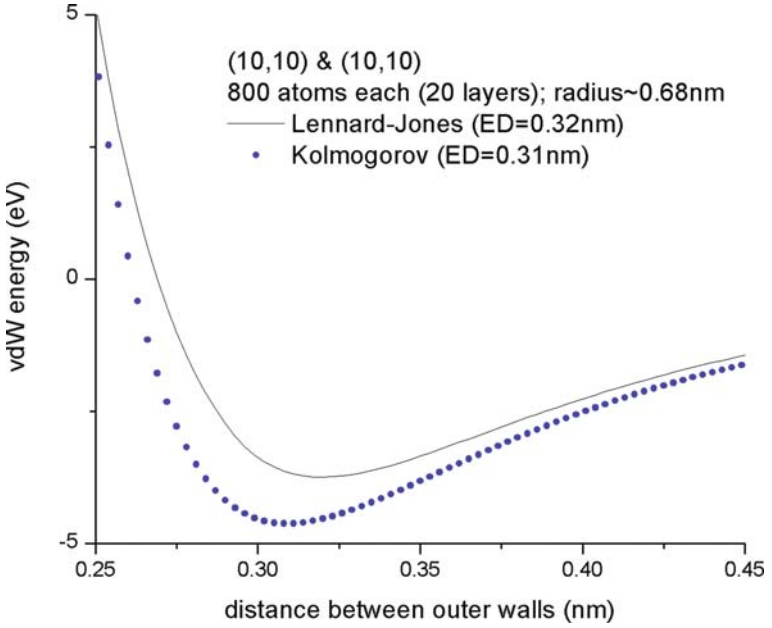


Figure 12-6. The van der Waals energy versus distance between two (10, 10) armchair single-wall carbon nanotubes given by Lennard-Jones potential and registry-dependent interlayer potential

sections of two carbon nanotubes after deformation induced by van der Waals interaction. The registry-dependent interlayer potential gives much larger deformation than the Lennard-Jones potential.

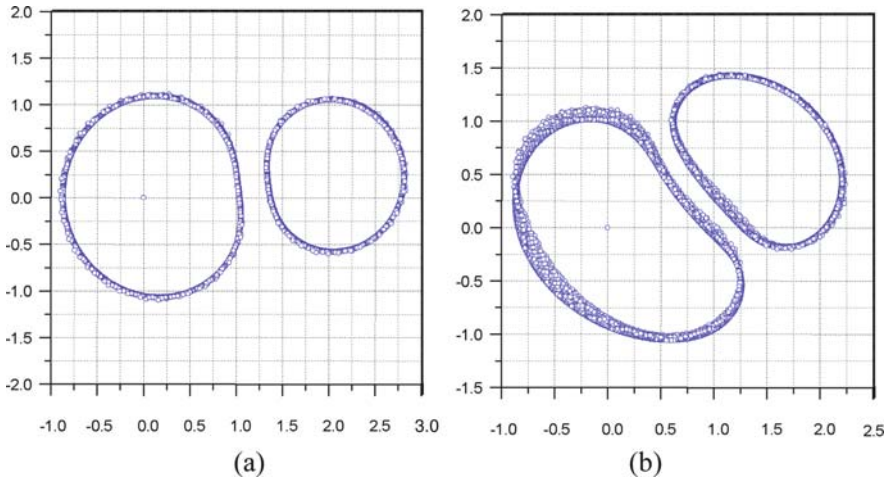


Figure 12-7. The cross sections of two chiral single-wall carbon nanotubes, (17,5) and (19,11), after deformation induced by van der Waals interaction; (**a**) the Lennard-Jones potential, and (**b**) the registry-dependent interlayer potential

The deformation of carbon nanotubes due to van der Waals interaction has been measured experimentally by the electron diffraction technique [22] through the intensity profiles of helical layer lines of the carbon nanotubes. Figure 12-8 shows the intensities of layer lines $(1\bar{1})$ and (10) for the (19, 11) tube and that of layer line $(1\bar{1})$ for the (17, 5) tube obtained from experiments by electron diffraction (grey crosses). [18] The intensity profiles obtained from the deformed carbon nanotubes simulated with the Lennard-Jones potential are shown in Figure 12-8 by blue dashed lines, while the black solid lines represent the intensities from the deformed carbon nanotubes with the radial deformation scaled by a factor of 0.75 which is obtained

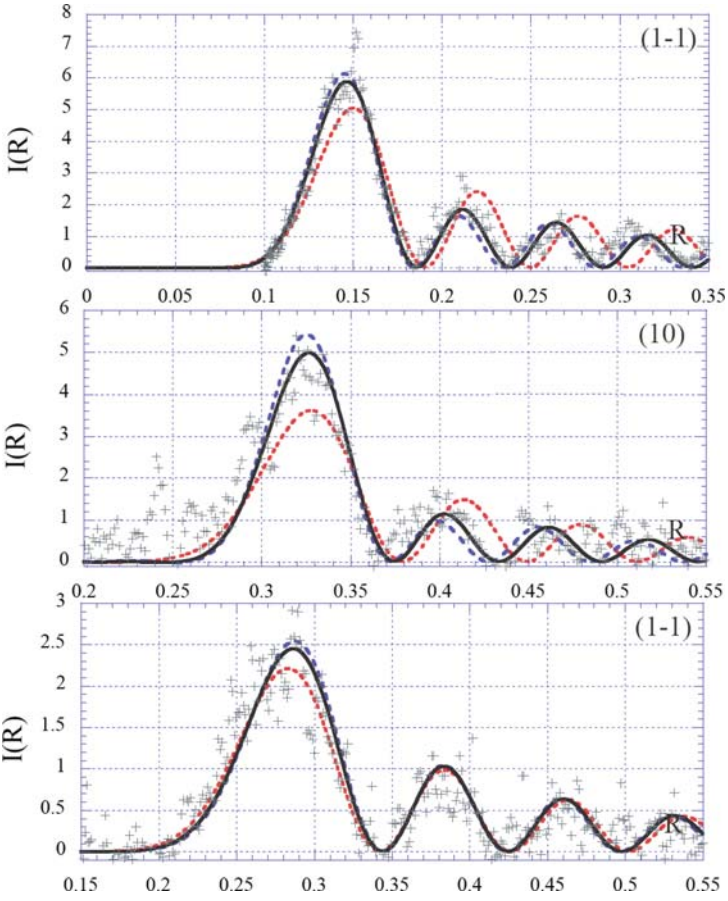


Figure 12-8. The intensities of layer lines $(1\bar{1})$ and (10) for the (19, 11) tube and that of layer line $(1\bar{1})$ for the (17, 5) tube obtained from experiments (*grey crosses*), undeformed circular carbon nanotubes (*red dotted lines*), deformed carbon nanotubes simulated with Lennard-Jones potential (*blue dashed lines*) and that with a scale factor 0.75 (*black solid lines*). Reprinted with permission from ref. 18. Copyright 2008 by the American Physical Society

from the fitting with experimental data. The intensities from the undeformed, circular carbon nanotubes are also given by red dotted lines for comparison. The ratio of the intensities of the second peak (I_2) over that of the first peak (I_1) could be used to quantify the agreement between simulations and experiments. For the $(1\bar{1})$ layer line for the (19, 11) carbon nanotube, the ratios are $I_2/I_1 = 0.19 \pm 0.02$ from experiments, 0.19 and 0.16 from simulations with and without 0.75 scale, respectively. The ratios are $I_2/I_1 = 0.21 \pm 0.04$ from experiments, 0.20 and 0.17 from simulations with and without 0.75 scale for the (17, 5) carbon nanotube, respectively. Though the 0.75 scaled radial deformation matches perfect with the experiments, the intensity profiles given by simulation with the Lennard-Jones potential without scale still agree reasonably well with the experimental data.

Table 12-1 compares the diameters of deformed carbon nanotubes from experiments and simulation. Here the diameter is the maximum dimension of the carbon nanotube in the direction perpendicular to the axis connecting the centers of two carbon nanotubes. The diameter given by the Lennard-Jones potential shows better agreement with the experiments than the registry-dependent interlayer potential.

Table 12-1. Comparison of the diameters of deformed carbon nanotubes from experiments and simulation

		Diameter from simulation (nm)	
Carbon nanotubes	Diameter from experiment (nm)	Lennard-Jones	Kolmogorov
(17,5)	1.60	1.63	1.77
(19,11)	2.25	2.16	2.48

12.5. CONCLUDING REMARKS

The Lennard-Jones potential and registry-dependent interlayer potential are compared through graphitic systems by AFEM. The Lennard-Jones potential has weaker lattice registry effect than the registry-dependent interlayer potential whose prediction agrees better with the first-principles calculations. On the other hand, the deformation of carbon nanotubes predicted by the Lennard-Jones potential shows better agreement with experiments than that by the registry-dependent interlayer potential.

REFERENCES

1. Chopra NG, Benedict LX, Crespi VH, Cohen ML, Louie SG, Zettl A (1995) Nature 377: 135
2. Yu MF, Dyer MJ, Chen J (2001) Phys Rev B 64: 241403(R)
3. Yu MF, Dyer MJ (2001) J Appl Phys 89: 4554
4. Yu MF, Kowalewski T, Ruoff RS (2001) Phys Rev Lett 86: 87
5. Liu B, Yu MF, Huang Y (2004) Phys Rev B 70: 161402(R)
6. Benedict LX, Chopra NG, Cohen ML, Zettl A, Louie SG, Crespi VH (1998) Chem Phys Lett 286: 490

7. Gao G, Cagin T, Goddard WA (1998) Nanotechnology 9: 184
8. Xiao J, Liu B, Huang Y, Zuo J, Hwang KC, Yu MF (2007) Nanotechnology 18: 395703
9. Park CJ, Kim YH, Chang KJ (1999) Phys Rev B 60: 10656
10. Martel R, Schmidt T, Shea HR, Hertel T, Avouris Ph (1998) Appl Phys Lett 73: 2447
11. Lu JQ, Wu J, Duan W, Liu F, Zhu BF, Gu BL (2003) Phys Rev Lett 90: 156601
12. Wu J, Zang J, Larade B, Guo H, Gong XG, Liu F (2004) Phys Rev B 69: 153406
13. Autumn K, Sitti M, Liang YA, Peattie AM, Hansen WR, Sponberg S, Kenny TW, Fearing R, Israelachvili JN, Full RJ (2002) Proc Natl Acad Sci USA 19: 12252
14. Ge L, Sethi S, Ci L, Ajayan PM, Dhinojwala A (2007) Proc Natl Acad Sci USA 104: 10792
15. Mahdavi A, Ferreira L, Sundback C, Nichol JW, Chan EP, Carter DJD, Bettinger CJ, Patanavanich S, Chignozha L, Ben-Joseph E, Galakatos A, Pryor H, Pomerantseva I, Masiakos PT, Faquin W, Zumbuehl A, Hong S, Borenstein J, Vacanti J, Langer R, Karp JM (2008) Proc Natl Acad Sci USA 105: 2307
16. Girifalco LA, Hodak M, Lee RS (2000) Phys Rev B 62: 13104
17. Kolmogorov AN, Crespi VH (2005) Phys Rev B 71: 235415
18. Jiang YY, Zhou W, Kim T, Huang Y, Zuo JM (2008) Phys Rev B 77: 153405
19. Liu B, Huang Y, Jiang H, Qu S, Hwang KC (2004) Comput Methods Appl Mech Eng 193: 1849
20. Liu B, Jiang H, Huang Y, Qu S, Yu MF, Hwang KC (2005) Phys Rev B 72: 035435
21. Brenner DW, Shenderova LA, Harrison JA, Stuart SJ, Ni B, Sinnott SB (2002) J Phys Condens Matter 14: 783
22. Gao M, Zuo JM, Twesten RD, Petrov I, Nagahara LA, Zhang R (2003) Appl Phys Lett 82: 2703

CHAPTER 13

ELECTRICAL CONDUCTION IN CARBON NANOTUBES UNDER MECHANICAL DEFORMATIONS

A. PANTANO

Dipartimento di Meccanica, Università degli Studi di Palermo, Viale delle Scienze – 90128, Palermo, Italy, e-mail: apantano@dima.unipa.it

Abstract: The enormous potential of carbon nanotubes (CNTs) as primary components in electronic devices and NEMS necessitates the understanding and predicting of the effects of mechanical deformation on electron transport in CNTs. In principle, detailed atomic/electronic calculations can provide both the deformed configuration and the resulting electrical transport behavior of the CNT. However, the computational expense of these simulations limits the size of the CNTs that can be studied with this technique and a direct analysis of CNTs of the dimension used in nano-electronic devices, particularly multi-wall CNTs (MWNTs), seems prohibitive at the present. Here a computationally effective mixed finite element/tight-binding (to be referred to as FE-TB) approach able to simulate the electromechanical behavior of CNTs devices is presented. The FE-based structural procedure computes the mechanical deformation of the CNTs and provides a tight-binding (TB) code with the atomic coordinates in the deformed configuration. The TB code is carefully designed to realize orders-of-magnitude reduction in computational time in calculating deformation-induced changes in electrical transport properties of the nanotubes. The FE-TB computational approach is validated in a simulation of laboratory experiments on a multiwall CNT and then used to demonstrate the role of the multiwall structure in providing robustness to conductivity in the event of imposed mechanical deformations

Keywords: Carbon nanotubes, Tight binding, Multiscale computation

13.1. INTRODUCTION

The sp^2 carbon-carbon bond in the basal plane of graphene is the stiffest and strongest in nature. Carbon nanotubes (CNTs), because of their cylindrical and nearly defect-free structures, may approach the maximum theoretical stiffness and tensile strength. Atomistic simulations have confirmed that CNTs should meet expectations for an extremely high modulus, however, experimental determination of mechanical behavior and properties have proved particularly challenging

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 335–365.
DOI 10.1007/978-1-4020-9785-0_13,

because of the small dimensions of the tubes, with diameters of tens of nanometers for MWNTs and about 1 nm for SWNTs, and lengths of the order of microns. Several experimental studies have attempted to measure the elastic modulus of carbon nanotubes by vibration of individual cantilevered CNTs [e.g. 1, 2] or by directly measuring the reaction force for an imposed displacement [e.g., 3, 4] and, by nanoindentation of vertically-aligned multi-walled carbon nanotube (MWNT) arrays. [5]

Most experimental measurements [e.g.,1–5] utilize beam theory to interpret the mechanical response of nanotubes and reduce data to an elastic modulus. Beam theory treats the nanotube as a homogenous solid tube, but does not take into account specific structural details of nanotubes, such as the nested tube structure of the MWNT and/or the single-atomic-layer nature of each tube wall. Given the apparent ambiguity in specifying a thickness for a single atomic layer, the radial wall thickness of a SWNT is usually taken to be $t = 0.340$ nm, the equilibrium spacing of two graphite layers. The thickness of an N-layered MWNT is analogously specified as $t = \mathrm{N}^{*}0.340$ nm. This value of t is used within beam theory to evaluate the cross-sectional area, A, and the area moment of inertia of the cross-section, I. Consistent with the concept established for graphite sheets, the representative wall thickness 0.34 nm of single-walled carbon nanotubes is associated with a Young's modulus of about 1.06 TPa (the in-plane stiffness of graphene).

Using electronic and/or molecular dynamics simulations of SWNTs, several investigators have found unloaded SWNTs to possess an internal (strain) energy per carbon atom that exhibits a $1/R^2$ dependence [e.g., 6–8], where R is the radius of the tube. This dependence suggests that the SWNT wall mechanically emulates a shell. These simulations, along with additional MD simulations of SWNT tension and compression [e.g., 6–9], determine a wall membrane stretching stiffness of $C = 56 \div 62$ eV/atom ($C = 58.2$ eV/atom for graphite) and a wall bending stiffness $D = 2.2 \div 3.8$ eV Å^2/atom. As a result, taking a wall thickness t of $t = 0.34$ nm, gives a wall modulus of $E = 1.06$ TPa (computed based on the membrane stretching stiffness); using the bending stiffness of shell theory, $D = (t^3E)/[12(1 - \nu^2)]$, this wall modulus and thickness pairing then produces a wall bending stiffness as much as 25 times greater than the actual wall bending stiffness of a SWNT. The discrepancy is due to the indicated ambiguity in specifying a thickness for the single-atomic-layer wall. Fitting aforementioned results from ab initio and semi-empirical MD studies [6, 10, 11] to elastic shell theory, it is possible to determine values of the effective Young's modulus and 'mechanical' wall thickness that enable shell theory to correctly predict both the wall membrane and wall bending stiffnesses of SWNTs. The surface Poisson ratio can be extracted from MD simulations to be ν=0.19, the same as the in-plane Poisson ratio of graphite. [12] Thus, the mechanical behavior of SWNTs can be captured by simply modeling the CNT wall as an elastic shell of effective mechanical thickness t_{wall}=0.075 nm and effective isotropic elastic modulus E_{wall}=4.84 TPa and Poisson ratio 0.19 as it has been demonstrated in Pantano et al. [13–15].

The importance of specific features of multi-wall CNT structure on its deformation behavior has been documented in several microscopy studies [e.g., 1, 16, 17].

MWNTs under bending develop reversible periodic wrinkles on the compressive side of the bend [e.g., 1, 16, 17]. Inter-wall spacing is preserved during buckling, demonstrating a mechanical effect of the strong normal van der Waals (vdW) interactions present between the walls of MWNTs. The wall-to-wall shear interaction is very compliant and weak [18] compared to the normal vdW stiffness and strength. Computational nonlinear structural mechanics studies of MWNT deformation have been presented by Pantano et al. [13–15] using the approach that will be described later in this chapter.

The effects of mechanical deformation on the electron transport behavior of carbon nanotubes (CNTs) are of primary interest due to the enormous potential of nanotubes in making electronic devices and nanoelectromechanical systems (NEMS). Significant changes in CNT conductivity are linked to structural features, including diameter, chirality and distortions [e.g., 19, 20], giving behavior ranging from narrow-gap or moderate-gap semiconducting to metallic. Conduction in defect-free CNTs has been observed to be ballistic in nature, implying the absence of inelastic scattering and involving little energy dissipation [e.g., 21]. Progress in theoretical understanding as well as experimental study and device realization in this field has been rapid [e.g., 21–44]; already several prototypical devices have been constructed and demonstrated [e.g., 45], including actuators, transistors and nano-switches. Most discussions of the electronic structure of CNTs assume perfect cylindrical symmetry, but this is somewhat of an oversimplification. High resolution images of CNTs often disclose structural deformations such as bent, twisted, or collapsed tubes. These deformations may develop during growth, deposition, and processing, or upon interaction with other CNTs, and with surfaces and surface features such as electrodes. Numerical simulations of the electromechanical behavior of CNTs mostly focus on SWNTs rather than on the more abundant MWNTs, and the few studies available in the literature deal with small segments of only two- or three-walled CNTs [40–43]. This is mainly due to the high computational burden involved in atomistic (i.e., ab initio, tight-binding, molecular dynamics) simulation of MWNTs of realistic dimensions. In order to enable and optimize CNTs as a basic building block of a new nano-electronic technology, a sufficient understanding of both the electrical and mechanical properties, as well as their dependence on mechanical deformations, must be achieved.

Experiments on the conductance of MWNTs, where the only the outermost wall is in contact with the electrodes and the inner walls being not in direct contact, have shown contrasting results on interwall conductivity. Indeed the majority of the experiments have revealed that current flows only in the outermost wall [24–29, 44], and there isn't any significant charge transfer across the shells. The main prove of this behavior is that the electrical conductivity does not scale with the number of walls in a MWNT [e.g., 24]. If the transmission coefficients are unity, at the Fermi level the conductance of individual undeformed metallic tubes has been predicted theoretically and verified experimentally [e.g., 21] to be exactly 2 G_0 , where the conductance quantum $G_0 = 2e^2/h \approx (12.9\ \mathrm{k}\ \Omega)^{-1}$. In the experiments of Frank et al. [24], where the tips of metallic MWNTs were immersed into liquid metal, a conductance of only 1 G_0 was observed for MWNTs with different diameters. Choi, et al.

[44] performed ab initio conductance calculations for a (10,10) CNT with one end immersed in a jellium metal trying to interpret results in [24], and finding that one of the two channels available for electron transport is turned off due to charge transfer between the SWNT and the jellium metal contact, creating a substantial band shift that reduced the conductance to 1 G_0. They [44] also consider the influence of the interwall coupling because experiments were done with multi-wall tubes and reach the following conclusion: the conductance in the MWNTs is determined by the outermost wall as if the system were a single wall nanotube. Using four-point measurement techniques Bourlon et al. [28] estimated the intershell electron transport in multiwalled carbon nanotubes. Measured intershell conductance was found in agreement with the estimate based on electrons tunneling through overlapping atomic π orbitals of nearby shells. Thus, the intershell conduction must differ from that in the sheet plane. Assuming a 0.34 nm intershell separation, they deduced a radial resistivity for MWNTs of about 1 Ω m, much larger than that of bulk graphite. Bourlon et al. [28] have also shown that the current flows mainly along individual shells that are rather efficiently insulated from each other. An interesting comment was given about deformed MWNTs, deformation can locally enhance or reduce the orbital overlap and thus affect the intershell conductance. Bachtold, Schönenberger et al. [26, 27] assembled a device where a single multi-walled carbon nanotube was electrically contacted with four electrodes. While imposing current I and measuring voltage V, the electrical resistances of single MWNTs were measured in a magnetic field B which was aligned parallel to the tube axis. They [26, 27] concluded that quantum-interference corrections to the resistance can account for the measured magnetoresistance, if and only if the electrical current is assumed to flow in one or at most two of the outermost shells. Tunney and Cooper [29] using a tight-binding formalism studied theoretically the electrical transport between shells in DWNTs. They have calculated that intertube conductance depends on the chirality of the two SWNTs making the DWNTs and can be classified in three different kinds: "zero", "intermediate", and "strong". But even in the "strong" case, intershell conductance was calculated to be smaller than $0.05\ e^2/h = 0.025\ G_0$; much smaller than the $2\ G_0$ conductance in the outermost wall of a MWNT which is undeformed and metallic. Their results also showed that localized depression of the outer shell can enhanced intertube conductance. In contrast with the studies presented earlier, Collins and Avouris [30] did show experimental results where there is significant contribution from inner shells to the overall conductance of the MWNTs. Their electronic device was made of MWNTs suspended among four to six gold electrodes. Their method consisted in a destructive removal of the outermost carbon shell from a MWNT. The method used current-induced oxidation that was selective because only the outermost shell was in contact with the oxidizing environment (air) at any given moment. As the outermost wall was destroyed the current drop of a small fraction and kept dropping as the following outer shells were eliminated, Collins and Avouris [30] concluded that the inner walls were contributing to the overall conductance of the MWNT. However, as the authors observed, this result, which conflicts with previous observations, may be justified by two factors. A high bias was used that determined multiple accessible conduction pathways for these inner

shells, including radial tunneling or charge injection directly from the electrode to a particular shell. Additionally, it is possible that the oxidation of an outer shell does alter the coupling between it and a neighboring inner shell. Moreover the measured nanotube resistances ranged between 5 and 15 kΩ, higher or slightly lower than the resistance of an ideal metallic nanotube, 6450 Ω; significant contribution to conductance from the inner walls should determine much smaller value of the resistance. Accordingly, it is reasonable to assume that only the outermost wall contributes to MWNT conductance. This finding allows for computational savings since the TB calculation for predicting the conductance need be performed on only the outermost wall of the MWNT. It is also worth to mention the work of Gupta et al. [25] as another experimental set up similar to the one that has been used to validate the proposed numerical approach. They used a nano-manipulation system operating in an SEM, which did also act as an electrical probing system, to characterize of individual carbon MWNTs. For an individual undeformed MWNT they measured a resistance of about 300 k Ω, which increases significantly upon bending of the nanotube. It was not possible to quantify the contribution of nanotube resistance and the contact resistance to the measured resistance.

Here we present a mixed finite element/tight-binding (FE-TB) approach able to simulate the electromechanical behavior of SWNTs and MWNTs of the dimensions used in nano-electronic devices. To realize the computational savings needed to work with realistic-sized CNTs, the mixed approach capitalizes on the ability and efficiency of the FE method to compute mechanical deformation of CNTs and the tight binding technique to compute electronic behavior if given atomic positions.

13.2. MODELING PROCEDURES

Elastic shell theory, formulated for structures with one dimension (the thickness) much smaller than the remaining two dimensions, appears a natural idealization for single-wall carbon nanotubes that can be thought of as one graphene layer of a hexagonal lattice structure that has been wrapped into a seamless cylinder. However, the applicability of continuum shell theories to the carbon nanotube structure, where walls are constructed of a single layer of atoms, is far from obvious. Satisfactory comparisons with MD simulations and experiments were needed to validate the accuracy of such an approach.

The experimental and modeling results reveal that a successful continuum-level model of SWNTs and MWNTs can be constructed by proper treatment of four key aspects of the CNT. First, the shell behavior of each carbon nanotube wall can be modeled using shell theory with an appropriate set of elastic constants and, as importantly, an appropriate corresponding *mechanical* thickness of the shell. Second, the tubular nature of the CNT structure necessitates a departure of the lattice from a stress-free planar hexagonal structure to a wrapped structure which thus contains an initial internal stress state corresponding to its initial curvature. Third, a CNT wall interacts with neighbouring walls, with other portions of itself, and with like substrates via strong van der Waals interactions. Fourth, the wall-to-wall shear

resistance is negligibly small. The manner in which each of these four items is captured in the proposed finite-element-based continuum-level representation of carbon nanotubes is discussed separately below.

13.2.1. The Carbon Nanotube Wall

Comparison of various stretching and bending behaviors of SWNTs to shell theory analysis leads to the determination of a membrane stretching stiffness C and a bending stiffness D of the wall, which, together, are used to compute an effective modulus and thickness pair (E_{wall}, t_{wall}) of the SWNT wall. The origins of the membrane stretching stiffness and the bending stiffness of the CNT walls can be described in terms of the atomic structure and properties of a graphene sheet. The hexagonal lattice of carbon atoms in a graphene sheet opposes deformation by resisting C–C bond stretching, in-plane C–C–C angle variation, dihedral angle torsion, and inversion (out-of-plane angle bending). The membrane stiffness of the nanotube shell is a measure of the resistance to in-plane stretching and results from the resistance to C–C bond stretching and the resistance to C–C–C (three-body) in-plane angle bending. The bending stiffness of the nanotube shell is a measure of the resistance to changes in the curvature of the shell. In macroscopic-scale shells, this stiffness arises from an integration of the distribution in tensile/compressive resistances through the thickness of the shell. In the nanotube, where the shell is of only single atomic layer thickness, the bending stiffness is a result of the resistance to the change in the C–C–C in-plane bond angle and dihedral angle torsion necessitated to accommodate the change in curvature of the atomic layer, as well as the inversion (out-of-plane bond angle variation) due to the overlapping of the electron clouds between atoms. Atomic-level models capture these interatomic interactions to various degrees and in different ways; therefore, each result in slightly different values for membrane and bending stiffness. Using a representative membrane stiffness of $C = 59.36$ eV/atom and a bending stiffness of $D = 2.886$ eV Å^2/atom, we can compute the (E_{wall}, t_{wall}) pair required to give the desired shell membrane and bending stiffnesses, giving E_{wall}=4.84 TPa and t_{wall}=0.075 nm; additionally, we took Poisson's ratio to be that of a graphene sheet, ν =0.19.

Often applications of traditional shell theories to CNTs found in the literature are restricted to infinitesimal deformations. However, post-buckling behaviors observed in both experiments and in atomistic simulations reveal important elastic behavior well beyond linear analysis. This predominantly small deformation elastic analysis contrasts with the fact that the nanotubes undergo very large nonlinear deformations elastically.

By modeling SWNTs using a continuum elastic shell model accounting for large geometric nonlinearities, it should be possible to appropriately capture their mechanical behavior. For simulations involving MWNTs, while the individual layers of the tube are modeled as SWNTs, the van der Waals interaction between walls needs to be implemented in its geometrically nonlinear form. Most commercial finite element codes contain shell elements able to correctly account for small strain/large rotations; however, among them the well-established code ABAQUS

was chosen for the investigation because of the presence of several programmable user subroutines which enable easy implementation of the van der Waals interaction. Thick shell theory is used as the shell thickness increases, yielding solutions for structures that are best modeled by shear flexible (Mindlin) shell theory. General-purpose shell elements become discrete Kirchhoff thin shell elements as the thickness decreases; the transverse shear deformation becomes very small as the shell thickness decreases. Experiments [e.g., 3], and atomistic simulations [e.g., 46, 47],have shown that CNTs can reach moderately large strain before defects are produced or fracture occurs; appropriately, adopted element types account for finite membrane strains.

The technique is further enhanced to account explicitly for the chirality of the CNT and to allow higher accuracy in modeling the deformation of small diameter SWNTs. Using quadrilateral shell elements, we construct a FE mesh of hexagonal cells in which all nodes correspond to individual atomic positions. The method enables FE modeling of CNTs of every type, including chiral ones. An example of the new meshing technique is provided in Figure 13-1, where a (13,0) SWNT ($D \cong$ 1.nm) is discretized by three superimposed meshes, Figure 13-1a–c, sharing all the nodes. Images of the final mesh are provided in Figure 13-1 d and e; in d, the sides of the shell elements are partially hidden for a clearer identification of the hexagonal cells. Successive rotation of shell element pairs within a hexagonal cell by 60° in the three superimposed FE models cancels the artificial mechanical anisotropy caused

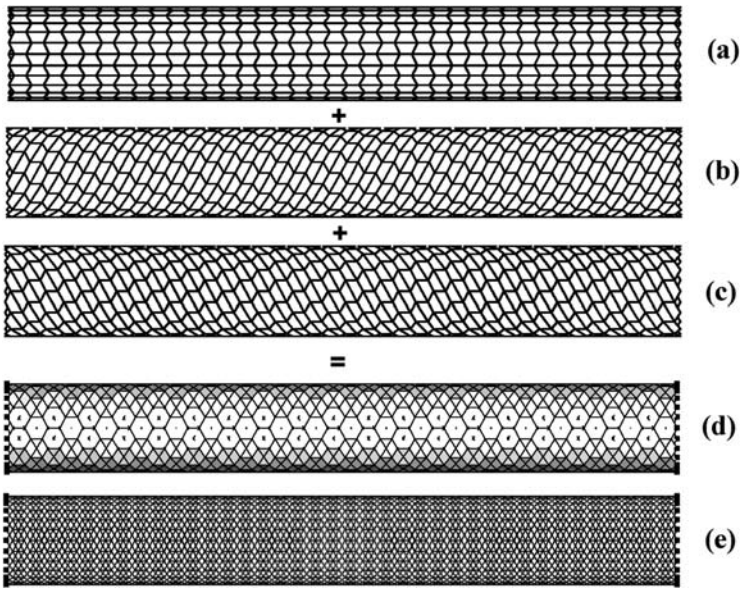


Figure 13-1. Construction of a FE mesh of a (13,0) SWNT composed of hexagonal cells, where all nodes correspond to individual atomic positions. The CNT is discretized by three superimposed meshes: (**a**)–(**c**). Images of the final mesh are provided in (**d**) and (**e**). In (**d**) the sides of the shell elements are partially hidden for a clearer identification of the hexagonal cells

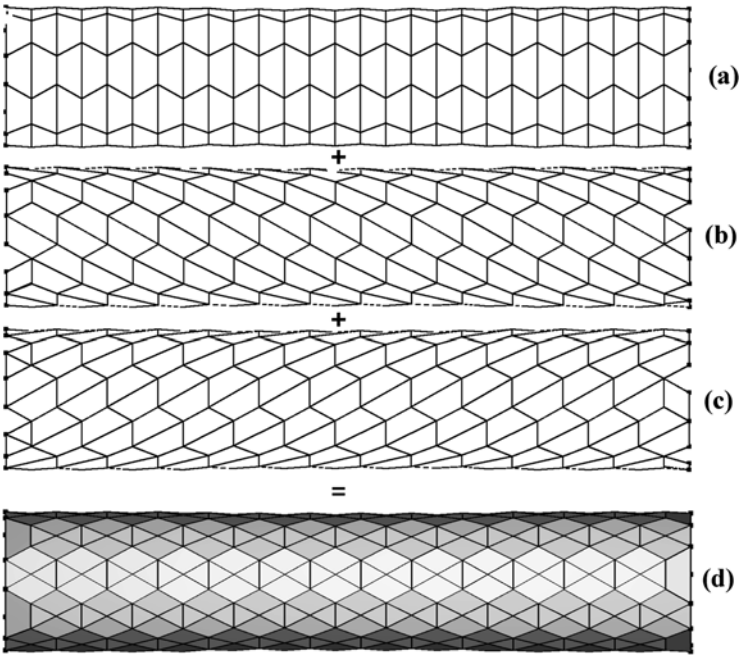


Figure 13-2. Construction of a FE mesh of a (5,5) SWNT. The CNT is discretized by three superimposed meshes: (**a**)–(**c**). Images of the final mesh are provided in (d)

by any single-orientation quadrilateral meshing of the cell. Figure 13-2 illustrates the mesh construction procedure for an armchair (5,5) nanotube. In order to obtain the proper overall CNT wall membrane stretching stiffness and bending stiffness, the Young's modulus assigned to each shell element is E_{wall}=1.6133 TPa (i.e., one-third of what should be assigned when a single-layer mesh is used to discretize the CNT wall); and the [unchanged] wall thickness of each element is specified to be t_{wall}=0.075 nm.

The CNT atomic positions correspond directly to nodal positions of the FE mesh. Therefore, in deformed configurations, atomic coordinates are precisely determined using the continuum/finite element approach, which then enables the effect of mechanical deformation on the electrical transport properties of nanotubes to be computed by a tight-binding (TB) method. In documentation of the commercial ABAQUS/Standard FE program (Abaqus Manuals), user subroutine URDFIL can be used to extract the nodal [atomic] coordinates at each time step of a mechanical simulation.

13.2.2. Initial Internal Stress State

Several authors [e.g., 6–8, 48] have demonstrated that carbon nanotubes contain internal stress due to their curvature. A continuum model should take the effect

of this initial stress state into account. The rolling energy of graphene is given by the following expression:

$$U_R = \frac{D}{2R^2} = \frac{1}{2} D\kappa_{\text{wall}}^2 = \frac{t_{\text{wall}}^3 E_{\text{wall}}}{24(1-\nu^2)} \kappa_{\text{wall}}^2. \tag{13-1}$$

Here R is the tube radius and $\kappa_{\text{wall}} = 1/R$ is the initial wall curvature; as before, E_{wall} is the Young's modulus of the wall and t_{wall} is the wall thickness.

When a finite element model of a carbon nanotube is developed, the initial internal stress state is computed by shell theory based on its radius R. The computed stress distribution through the shell thickness is then applied as a pre-existing stress state to the model. For large-diameter MWNTs, the influence of the initial stress state is negligible, since as the tube diameter grows, the initial stress values diminish rapidly. However, the subsequent deformation of small-diameter SWNTs can be considerably influenced by the presence of the initial [bending] stress state.

13.2.3. Construction of Special Interaction Elements

The development of an interaction element able to simulate the van der Waals forces is necessary in order to study MWNT deformations, post-buckling behavior of SWNTs, and tube/tube and tube/substrate interactions.

Two major functional forms have been used in empirical models of the van der Waals interaction potential: the inverse power model and the Morse function model. A very widely used inverse power model is the Lennard-Jones (LJ) potential. For the carbon-carbon system, the atom/atom non-bonded LJ potential energy has been treated by Girifalco et al. [49, 50] and is given as:

$$\phi_i = \frac{A}{\sigma^6} \left[\frac{1}{2} y_0^6 \frac{1}{\left(\frac{r_i}{\sigma}\right)^{12}} - \frac{1}{\left(\frac{r_i}{\sigma}\right)^6} \right]. \tag{13-2}$$

Here σ=0.142 nm is the C–C bond length, y_0 is a dimensionless constant, and r_i is the distance between the ith atom pair. For graphite, Girifalco and Lad [50] determined A=24.3·10^{-79} J m^6 and $y_0 = 2.7$. In the literature [e.g., 51–55] the LJ and local density approximation (LDA) potentials are compared by computing the equation of state (EOS) for a graphite system by assuming no relaxation within the graphene plane. Two graphene sheets (in AA and AB registry, respectively) are considered. The in-plane geometry is held fixed, and the inter-layer distance is changed. The computed pressure, as a function of the volume change, is plotted. The results are compared with published experimental data by Zhao and Spain [51] and by Hanfland et al., [53] and with the ab initio treatment by Boettger. [55] Both models provided good fits to the experimental data and to Boettger's model, but the LDA model seems too attractive at distances greater than 0.340 nm. The resulting expression for the pressure as a function of the inter-layer distance computed from

the Lennard-Jones (LJ) potential has been implemented in an ABAQUS user subroutine for simulating the van der Waals forces. The pressure/inter-layer-distance relation calculated by Zhao and Spain [51] has been adopted:

$$p(\alpha) = \frac{\Psi}{6}\left[\left(\frac{d_0}{\alpha}\right)^{10} - \left(\frac{d_0}{\alpha}\right)^{4}\right]. \tag{13-3}$$

Here p is the pressure, α is inter-layer distance, $d_0 = 0.340$ nm is the equilibrium distance, and Ψ=36.5 GPa.

In ABAQUS/Standard (ABAQUS Manuals) user subroutine UINTER can be used to define the constitutive interaction between two deforming surfaces. The interface is taken to be massless. The developed interaction elements are able to accurately predict the pressure/inter-layer-distance variation between two planar graphene sheets, since Eq. (13-3) has been verified to agree with experimental results. [51, 55] To test the validity of the proposed approach, three different kinds of problems have been simulated and compared to MD results: [13] self collapse of SWNTs, interaction of SWNTs with a substrate, and tube-tube-substrate interaction. In all of these analyses the interaction elements have been able to simulate successfully the van der Waals forces, demonstrating that, even in case of surfaces with different radii of curvature, accuracy of the results is not drastically affected. The reason for the limited sensitivity to the radius of curvature recorded can be found in the rapid distance decay of the van der Waals forces with the atom/atom distance. The force between two atoms as a function of the inter-atomic distance can be derived from the expression for the potential, Eq. (13-3), by taking the derivative with respect to the distance. It follows a (13,7) dependence on inter-atomic distance, resulting in forces that are already negligible at distances as small as $2d_0 = 0.68$ nm, about twice the equilibrium interplanar distance.

13.2.4. Model of the Inter-Layer Shear Resistance

As discussed above, in contrast to the stiff and strong normal wall-to-wall interaction, there is relatively low shear stiffness and strength between layers in MWNTs. In fact, the shear strength has been found to be less than 0.5 MPa, the mean value of the inter-layer sliding resistance strength for crystalline graphite reported in the literature. [12] Thus, it seems a reasonable assumption to model the shear stiffness and strength to be zero, as was also done by Ru [56] in his continuum study on the effect of van der Waals forces on axial buckling of a double-walled carbon nanotube.

13.2.5. Electrical Transport Model

Within the FE program simulating the mechanical deformation of the nanotube structure, the evolving atomic [nodal] coordinates are further processed using a tight-binding (TB) code which calculates deformation-induced changes in electrical

transport properties of the nanotube. There the coordinates are used in TB calculations to compute electronic properties of the system. Our main objective is to predict effects of deformation on the electrical properties of MWNTs; hence, adopting a conventional TB code would negate the computational saving obtained by using the structural mechanics approach to compute the deformed atomic coordinates. Instead, a new TB algorithm was developed that results in substantial computational savings. Beginning with a recently-developed approach, the coherent transport properties of infinitely-long or finite conductors spanning the distance between two leads can be computed using two different full sp^3, four-orbital, orthogonal TB models [57, 58] or a non-orthogonal TB model. [59] For details on the quantum conductance calculations, the reader is referred to. [37] Here, the full sp^3, four-orbital, orthogonal TB model [58] has been used. Then starting from the methodology for computing the conductance of a system with a large number of atoms [60], a novel approach has been developed where each CNT is divided along its length into a number of shorter CNT segments connected to one another at interfaces. The system is equivalent to the original CNT, but the memory requirements and the computational expenses are reduced by orders of magnitudes. The realized computational efficiencies increase with increasing dimensions of the nanotube; for example, the analysis of a (70,70) tube had a memory requirement almost 5 orders of magnitude smaller than a conventional coding implementation of the TB algorithm. The portion of the (70,70) tube that is allowed to deform was divided along its length in 207 pieces, since the matrices on which the program operates are square with dimensions equal to [(number of atoms) $\times$ (number of orbitals per atom)]^2, the reduction in the memory requirement is equal to $207 \times 207 = 42849$, which gives a ratio of 0.000023 with respect to the requirements of the original method.

13.3. NUMERICAL RESULTS

13.3.1. Bending of SWNTs

Here, the proposed FE-based mechanical approach was first validated by comparison of simulation results with the MD results of Yakobson et al. [9] for axial compression and bending of SWNTs of various diameters. Our numerical results for strain energy and critical buckling loads over a wide range in tube geometries compared very favorably with the MD results.

In particular, elastic strain energy prior to buckling was within 1.5% of the MD results. Here, an exemplar bending case is chosen for presentation in order to better demonstrate the agreement between the new FE-based simulation approach and MD results. Figure 13-3 shows the strain energy during bending of a (13,0) helicity ($R \cong 1.0$ nm) and 8 nm long SWNT as a function of bending angle, along with Yakobson's MD results, [9] for various model parameters. Bending was imposed by rigidly rotating the planes of nodes of the two ends of the originally straight tube through equal and opposite angles. The axis of rotation passed through the tube's

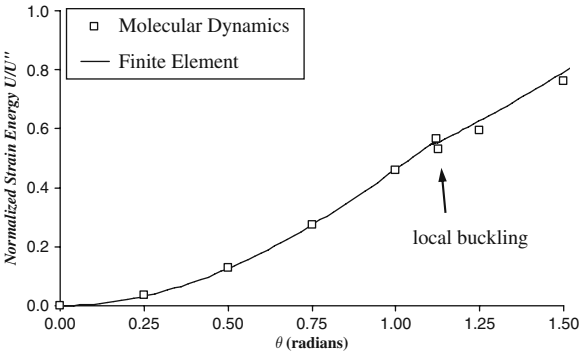


Figure 13-3. Comparison of the total strain energy $U(\theta)$ (normalized by its initial second derivative, $U''(\theta = 0)$) as a function of bending angle θ for an 8 nm long nanotube of (13,0) helicity ($R \cong 1.0$ nm)

center and was normal to the (unbent) length of the tube. Figure 13-3 plots the total energy of imposed bending, $U(\theta)$, normalized to its second derivative at $\theta = 0$, as a function of increasing bending angle. Prior to buckling, the differences in energy between the computational approaches are negligible, and the curves exhibit an essentially perfect quadratic behavior. The buckling point predicted by the finite element simulation agrees well with the MD result. After buckling, differences among the solutions remain small.

13.3.2. Tube-Tube-Substrate Interaction

Hertel et al. [61] found that the van der Waals interaction between nanotubes, as well as between nanotubes and a substrate, can lead to substantial axial and radial deformations of adsorbed nanotubes. They dispersed multi-walled carbon nanotubes on an H-passivated Si(100) substrate. The nanotubes were imaged in air with an atomic force microscope operating in the noncontact mode. When CNTs cross obstacles such as other CNTs on the surface, they bend and deform elastically. The observed flattening of adsorbed nanotubes on the side facing the substrate leads to an increase in the contact area and binding energy with respect to undeformed tubes. Hertel et al. [61] also simulated two crossing (10,10) nanotubes on a graphite surface using molecular mechanics. They computed a tube-tube contact force about 5.5 nN, leading to cross-sectional distortion of both tubes.

This problem appears to be a valuable test for the present finite element approach, since it requires both correct modeling of the van der Waals interactions and a realistic structural response for the two small-diameter SWNTs. The equilibrium configuration for the two (10,10) nanotubes crossing at 30° angle on a graphite substrate, calculated by the finite element model, is shown in Figure 13-4.

Not only does it compare well with the final deformed configuration from the molecular dynamics analysis, but it predicts the contact force between the upper and the lower tubes to be 5.56 nN, almost exactly the same as the 5.5 nN found by Hertel et al. [61].

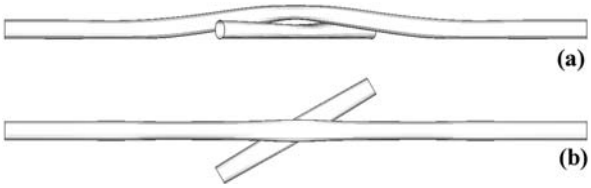


Figure 13-4. Equilibrium configuration for two crossing (10,10) nanotubes on a graphite substrate calculated by a finite element model. (**a**) Lateral view and (**b**) top view of the equilibrium configuration

13.3.3. Deformation of MWNTs Under Bending

Due to the prohibitive computational cost of the MD techniques when used to simulate deformation of carbon nanotubes of more than a few walls, energy/force/moment versus displacement results are generally lacking in the literature. Our attention thus focused on comparing high-resolution images of bent MWNTs with our finite element predictions.

Bower et al. [16] investigated by TEM composites of uniaxially oriented multi-walled carbon nanotubes embedded in polymer matrices. In strained composite films, buckling was observed in MWNTs bent to large curvatures. Over thirty such nanotubes with different inner and outer diameters were observed in TEM and analyzed. The buckling wavelengths were not accurately predicted by thin shell theory, [62] which predicts the buckling wavelength, λ, for a circular cylindrical shell under axial compression to be:

$$\lambda = \frac{2\pi}{\sqrt[4]{12(1-\nu^2)}}\sqrt{R_e h} \tag{13-4}$$

where R_e and $h = R_e – R_i$ are the outermost radius and the total thickness of the nanotube, respectively. If the Poisson's ratio is taken to be 0.19, then $\lambda = 3.4\ (R_e h)^{1/2}$. This is not in accord with experimental observations, where the experimental ratio $\lambda/(R_e h)^{1/2}$ has been found to be about 1. However, the buckling wavelength of SWNTs or MWNTs with very few walls agreed with Eq. (13-4). As previously stated, the underlying reason for the lack of agreement between the experimental results [e.g., 16, 63] and shell theory is that, since traditional elastic shell models are mainly directed to thin single-layer shells, they cannot be applied directly to carbon MWNTs, due to their multi-layer structure and the associated inter-wall van der Waals forces. In contrast, the SWNT can be considered as a thin-walled cylinder, and shell theory can well predict its behavior. The local buckling in the MWNTs is a complicated form of buckling, where van der Waals interactions with neighboring tube walls provide a constraint against the buckling analogous to an elastic foundation.

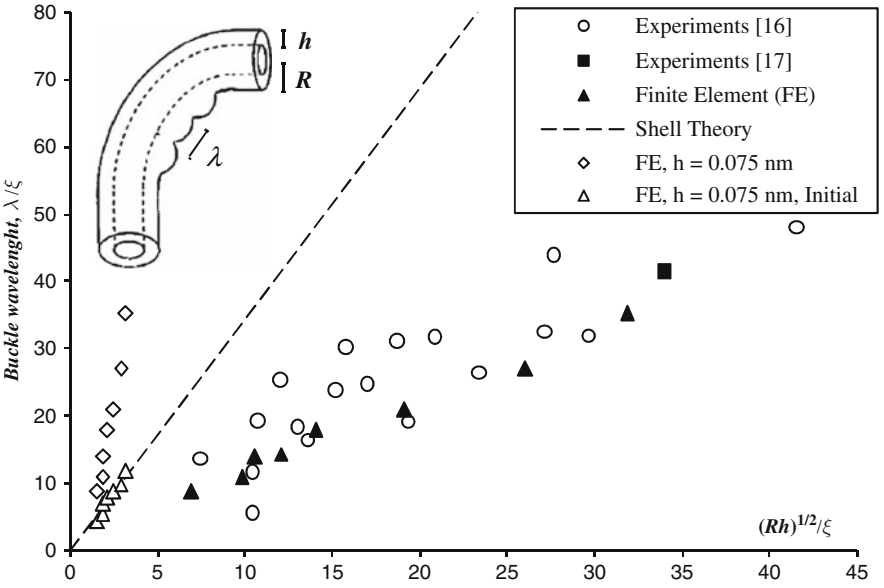


Figure 13-5. Normalized axial buckling wavelengths, λ, of the buckled MWNTs vs. a normalized geometric dimension of the nanotubes. Here $\xi = 0.24$ nm, R is the outer radius and h is either total MWNT thickness or SWNT effective thickness 0.075 nm [*open diamonds* and *open triangles*]

Figure 13-5 plots various model (shell theory and the new FE-based model) and experimental results of λ vs $(Rh)^{1/2}$, where both lengths have been normalized by $\xi = 0.24$ nm, the height of the hexagonal graphene lattice cell. As shown in Figure 13-5, taking h to be the thickness of the outer wall (t_{wall}), the *initial* λ computed by the new FE-based approach (open triangles) is in excellent agreement with thin shell theory (which considers only a single thin tube). As discussed earlier, the final λ greatly exceeds its initial value for all cases; this results in the much steeper-sloped line shown in Figure 13-5 (open diamonds). Figure 13-5 shows the experimentally-observed wavelengths, where Bower (open circles) took h to be the total tube thickness; additional experimental results [17] are plotted with filled squares. In Figure 13-5, final λ values computed in the FE-based simulations (filled triangles) are again reported, but this time the real multi-layer structure is assumed in modeling the MWNTs. The agreement of the new computational results with experiments is remarkable.

When MWNTs are bent, they develop a characteristic wavelike distortion or ripple on the area under compression. Their large radius makes the outer walls more compliant in the radial direction under bending. As the radius of the MWNT increases, the compressive strain increases for the same bending angle. The "rippling" mode is mainly dominated by partial release of a buildup in compressive stress. Once the outer layer buckles, the van der Waals interaction transmits loads to the inner walls, thus opposing abrupt large deformations of the outer wall.

Therefore, two main factors determine the buckled pattern: the bending stiffness of the walls and the van der Waals forces. In all of the experimental observations [e.g., 16, 17] the amplitude of the ripple in the large-diameter MWNTs decreased monotonically from the outside wall to the inner walls. There were no discontinuities in consecutive inter-layer spacings, nor was there evidence of defects. A direct comparison between a TEM image of a buckled MWNT and a corresponding finite element solution shows that not only do the buckling wavelengths closely match, but also the overall deformed configurations are correctly predicted. In their work Bower et al. show a TEM image, [16 Figure 2b] of a buckled MWNT in a CNT-polymer composite. We simulated the bending of a MWNT with the same geometry, same outer and inner diameter and same numbers of walls.

According to the TEM image shown in, [16] we modeled the MWNT as a 14-walled CNT constructed of nested (n,n) tubes, with n = 5,10,15,. . .70; the outer tube radius is R=4.76 nm. The length of the tube is L=15 nm. A bending angle of θ = 0.4 radians is gradually applied, imposing global curvature $\kappa = \theta/L$; the maximum κ is 0.0133 (1/nm). At one end, each tube section is rigidly rotated while radial and circumferential displacements are constrained; the opposite tube end is a symmetry plane. The total moment vs. global curvature diagram is shown in Figure 13-6.

Deformed configurations and contours of the inter-wall pressure for the 14-walled CNT are shown in Figure 13-7 at selected imposed curvatures, elucidating inter-wall interactions and the progression of wrinkling. The first deformed configuration, marked a, corresponds to the onset of buckling; its location on the moment-curvature diagram is indicated in Figure 13-6. For any given κ, the peak compressive strain in successive walls increases linearly with respect to wall radius. The outer walls cooperatively buckle at a κ value when local wall bending and vdW potential storage becomes energetically more favorable than local compressive stretching. Initial buckling occurs with a relatively small axial wavelength, λ; as the outer layers buckle, vdW interactions develop, opposing relative radial motion of the outer walls. As κ increases, the outer tubes attempt to increase their wrinkling

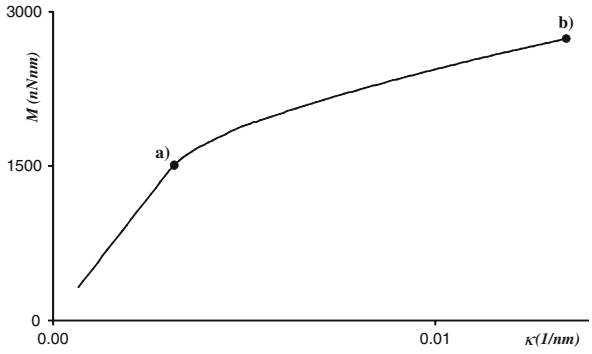


Figure 13-6. Pure bending of 14-walled carbon nanotubes constructed of nested (n,n) *tubes*: Total moment, the sum of the moments acting on all 14 tube walls, vs. global curvature

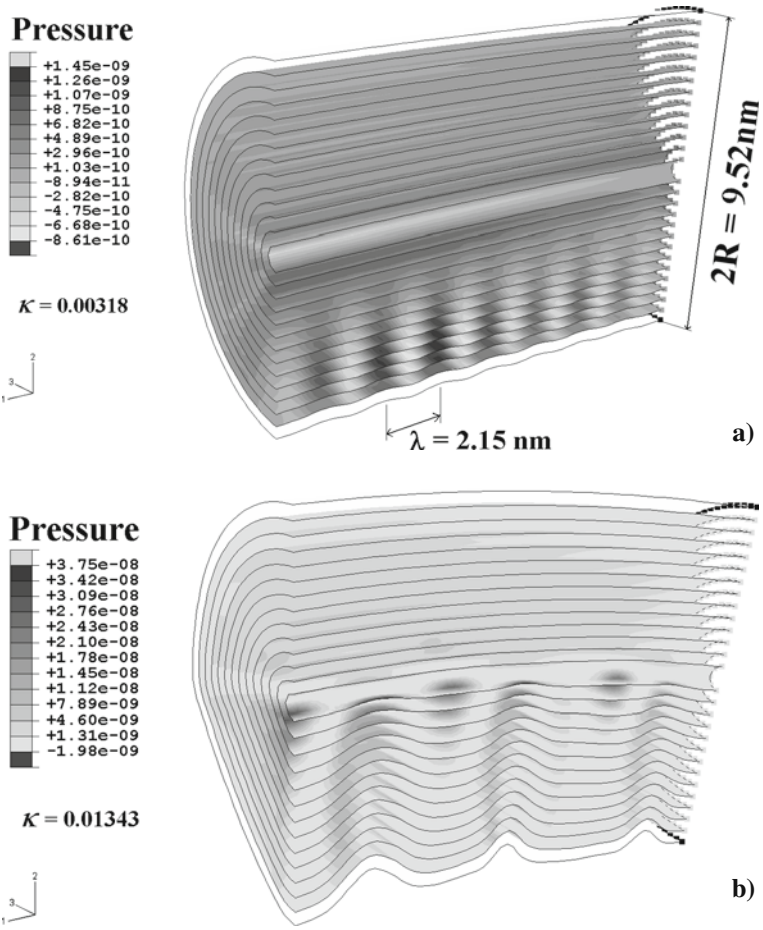


Figure 13-7. FE bending simulation of a 14-walled MWNT model at indicated curvatures, κ, in 1/nm. Distribution of the inter-wall pressure in N/nm^2 at the locations (**a**), and (**b**) of Figure 13-6

amplitude, but the vdW forces resist the developing inter-wall radial separation; instead, the wavelength λ increases with a further increase in global curvature, ultimately resulting in a buckling wavelength much longer than that initially observed at buckling onset.

Furthermore, the wavelength evolution is associated with a sharpening of the intrusions, as is also observed experimentally. [1, 16] This sharpening results from the nonlinear and directional nature of the vdW interactions. As κ increases further (Figure 13-7b), the wrinkle permeates to the innermost wall; no noticeable changes in inter-wall spacings are observed. The 14-walled CNT simulation compares well with the TEM image of Bower et al. [16, Figure 13-2b] in terms of both final λ and deformed configuration. The same progression is observed in all of the MWNT cases simulated. In our simulations the local radius of curvature at the deepest kink

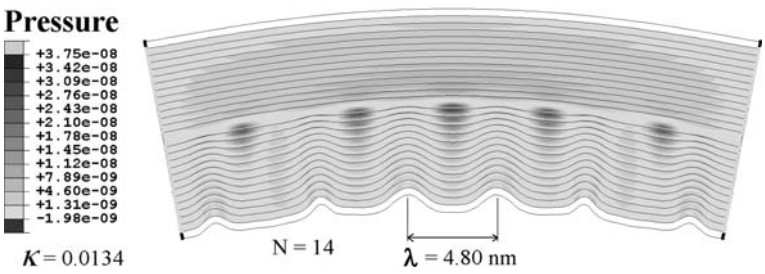


Figure 13-8. FE bending simulation of a 14-wall MWNT ($2R_e$=9.52 nm, $2L$=30 nm) – Axial section. Deformed configuration and distribution of the inter-wall pressure, in N/nm^2, at the final imposed curvature, in 1/nm, are shown

is never smaller than 0.27 nm. Thus, no complete and irreversible sp^2 to sp^3 transition should take place. [64, 65] Figure 13-8 shows the final deformed configurations for the 14-walled CNT discussed. Depiction of axial sections of the models and their images reflected across the symmetry plane provides a clearer view of the characteristic wavelike distortion on the area under compression; for the 14-walled tube, the final axial wavelength in the central section is λ=4.80 nm, considerably larger than the initial value of 2.15 nm shown in Figure 13-7a.

13.3.4. Laterally-Squeezed (8, 8) SWNT

In Lu et al. [35], TB molecular dynamics and the Green's function method are used to demonstrate that lateral squeezing of an armchair SWNT can induce a metal-to-semiconductor transition (MST). They employed a four-orbital TB approach for investigations of both the mechanical analyses and the electronic transport properties. Here, we utilize our electro-mechanical approach to reproduce the Lu et al. [35], simulations of the effects of squeezing on the electrical properties of an (8,8) armchair SWNT. A schematic of the geometry and loading configuration is provided in Figure 13-9; for details on the Lu et al., simulations, see [35].

The SWNT is laterally squeezed between two identical rigid tips, each with a width of 0.580 nm, initially separated by 1.087 nm (the initial diameter of the SWNT). It is well known that a perfect (8,8) armchair SWNT exhibits metallic behavior, as also found in the TB calculations of Figure 13-10. When the separation between the two tips is reduced to 0.700 nm, there is a very small change in the conductance.

When the separation is further reduced to 0.260 nm, in addition to the strong change in the curvature of the walls, the orbitals of atoms initially on opposite sides of a horizontal diameter start to overlap. At this level of deformation, a deformed configuration preserving initial mirror symmetry with respect to the compression axis exhibits electronic behavior different from that of one lacking such mirror symmetry of atomic positions. To show this dependence on symmetry, two cases are investigated: in one case, atomic positions retain mirror symmetry about the tip

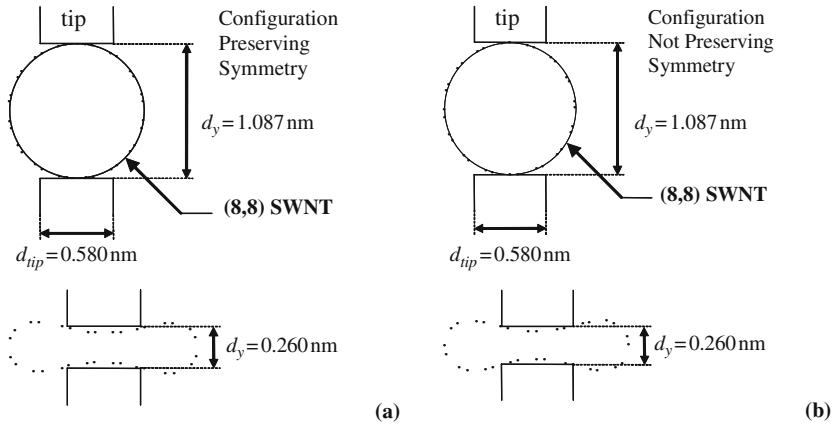


Figure 13-9. Initial and deformed configurations of an armchair (8,8) SWNT squeezed between two tips: (**a**) configuration preserving vertical mirror symmetry; (**b**) configuration not preserving vertical mirror symmetry

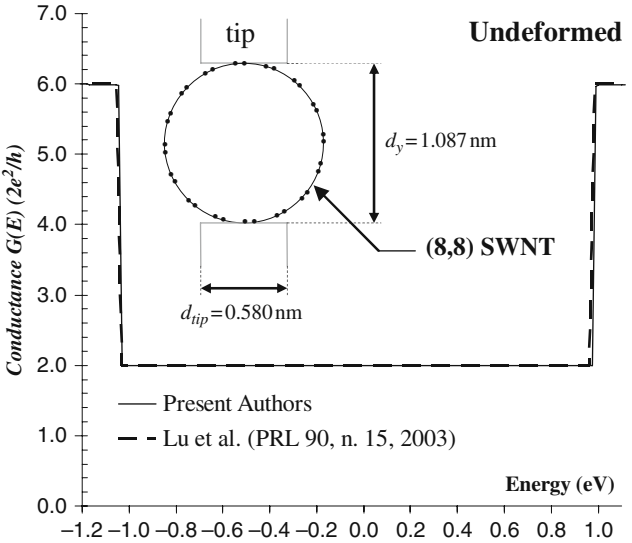


Figure 13-10. Conductance of an armchair (8,8) SWNT squeezed between two tips in the undeformed configuration

centerline (the compression axis), while in the other, initial symmetry is broken by a rotation of the SWNT by 7.75 degrees, corresponding to Lu et al. [35], Figure 1c and e, respectively]. The effect on the conductance is notable, as shown in Figures 13-11a and b: for a configuration preserving mirror symmetry, Figure 13-11a, the CNT retains its metallic behavior, whereas in the configuration breaking symmetry, Figure 13-11b, a band gap develops. The charge density plots reported in Lu, et al.

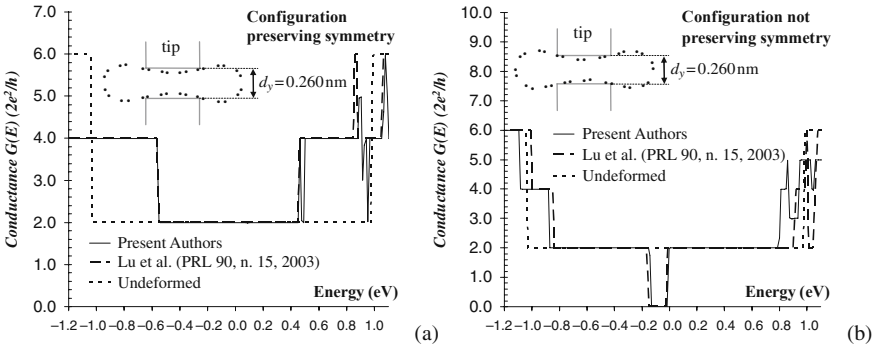


Figure 13-11. Conductance of a laterally-squeezed armchair (8,8) SWNT when the two tips have a distance of 0.260 nm and: (**a**) the deformed configuration preserves the initial mirror symmetry with respect to the direction of compression; (**b**) the deformed configuration does not possess mirror symmetry, owing to the rotation of the initial configuration

[35, Figure 2] further emphasize these differences. It should be underlined that, with further squeezing, both tubes become semiconducting. Also, good agreement between our calculations and those from Lu et al. [35] is evident in Figure 13-11.

13.3.5. Bent (10, 0) SWNT

Let's now focus on the effect of a bending deformation on the electronic transport properties of a (10,0) zig-zag SWNT. The nanotube is 9.8 nm long and is bent by fixing the last 8 carbon rings (80 atoms) at each end of the CNT, and then gradually rotating these rigid portions of the tube. The simulation replicates the study of Farajian, et al. [36], where a four orbital-per-atom tight-binding approach is used both to obtain the optimized geometries and to calculate the electronic and transport properties.

The CNT locally buckles at an imposed relative rotation near 120°, developing two clear kinks near the non-deforming ends at ∼130°, in good agreement with the calculations of Farajian et al. [36]. The results of the mechanical analysis in [36] show that this magnitude of bending deformation does not result in bond breaking or the formation of lattice defects in the nanotube, nor does it result in a collapse from graphite (sp^2) to diamond (sp^3) structure; thus the intrinsic stretching and bending stiffnesses of the wall should be largely unchanged, as is implicitly assumed in our de-coupled FE-based mechanical analysis.

The calculation of the electrical properties of the CNT is of the finite-length type, with undeformed leads at the ends. The leads are specified to be ideal nanotubes, constituting the rigid regions in the mechanical analysis. Figure 13-12 shows our calculations of the conductance for the (10,0) SWNT in the undeformed configuration and in the final 180° bent configuration. It is known that an undeformed (10,0) SWNT is semiconducting and, furthermore, no transition from semiconductor to metallic can take place since, regardless of what happens in the deformed regions,

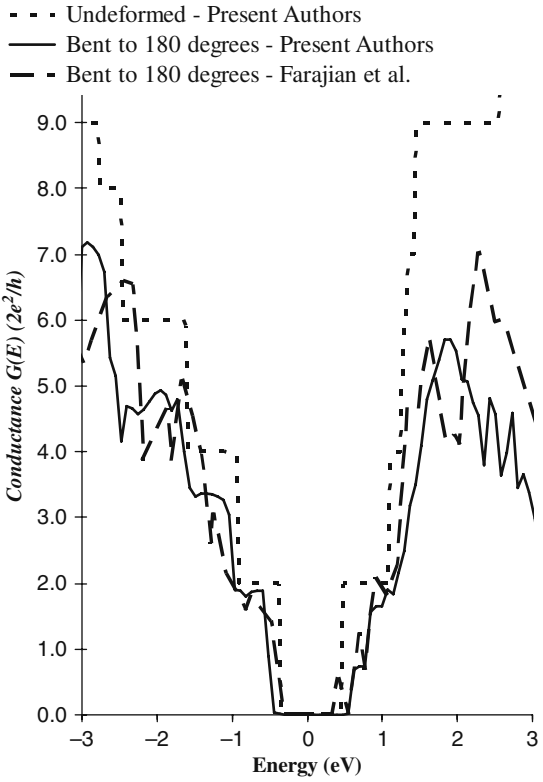


Figure 13-12. Conductance of a symmetrically-bent (10,0) SWNT. Different curves correspond to: undeformed configuration, and deformed configurations at 180° imposed rotation, from both the present formulation and from the results of Farajian et al. [36]

the undeformed leads remain semiconducting. Comparing our results with the calculations of Farajian, et al. [36], a good level of agreement is observed. Considering the significant differences in the parameterizations of the mechanical approach and in the TB codes, this agreement between outcomes provides a strong validation of the proposed FE-TB approach.

13.3.6. Simulation of Laboratory Experiments on a MWNT

Here we utilize our electro-mechanical modelling technique to reproduce the MWNT laboratory experiments of Kuzumaki and Mitsuda [66], who evaluated deformation-induced changes in electrical conductivity of a multiwall carbon nanotube. They measured electrical conductance during the deformation of MWNTs in a transmission electron microscope (TEM). Using a nanoprobe manipulation unit consisting of a fixed stage and a piezo-driven stage fitted into a TEM holder, they applied a compressive axial load on a MWNT, along with a bias voltage between

the two stages. A combination of axial compression and bending deformation observable in the TEM was induced in the CNT. For more details about the experimental setup and the testing methodology, the reader is referred to [66]. The applied bias voltage between tips was fixed at 1 V while the current through the MWNT was measured during deformation, starting from the undeformed configuration. In the undeformed configuration, the current measured through the nanotube exceeded 115 μA under an applied bias voltage of 1 V; the resistance of the MWNT obtained is R =V/I = 8696 Ω. This value is not far from the resistance of an ideal metallic nanotube, 6450 Ω [e.g., 2], indicating that the contact resistance is limited and not predominant as in other experiments available in literature. Since the undeformed conductance is near the theoretical one, we take the outermost CNT to be an armchair metallic (70,70) nanotube in the FE-TB model. Kuzumaki and Mitsuda [66] recorded a 10–12% reduction in the conductance when the MWNT was deformed to the level shown in their Figure 2b [66]. As the authors noted, the MWNT was not damaged (e.g., no Stone-Wales transition) at the imposed deformation level: the local radius of curvature at the deepest kink of the two buckling units formed was not small enough to cause a complete and irreversible sp^2 to sp^3 transition [64, 65]. This was confirmed by measurements taken after removal of the load: the nanotube recovered its original undeformed shape, and the current returned to its initial level.

As previously motivated, the outermost wall is assumed to be an armchair CNT, thus showing metallic behavior. The model is 60 nm long, and the diameter of the outermost tube is 9.492 nm. It is compressed and bent by fixing the atoms of each wall near the two ends, and then displacing these 4 nm long rigid portions; thus, the deformable part of the model is 52 nm long. Figure 13-13 shows the FE model in its deformed configuration, both a side view of the outer surface and a diametral sectional view, with an inset illustrating the hexagonal meshing of the tube. The model reproducing the MWNT used in the experiment was gradually deformed by

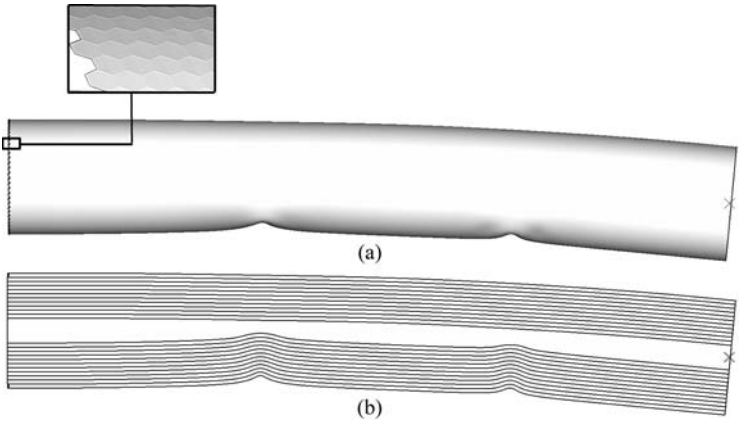


Figure 13-13. Final deformed configuration of the FE model reproducing the experiment, shown both in (**a**) a side view of the entire MWNT and (**b**) in a diametral sectional view

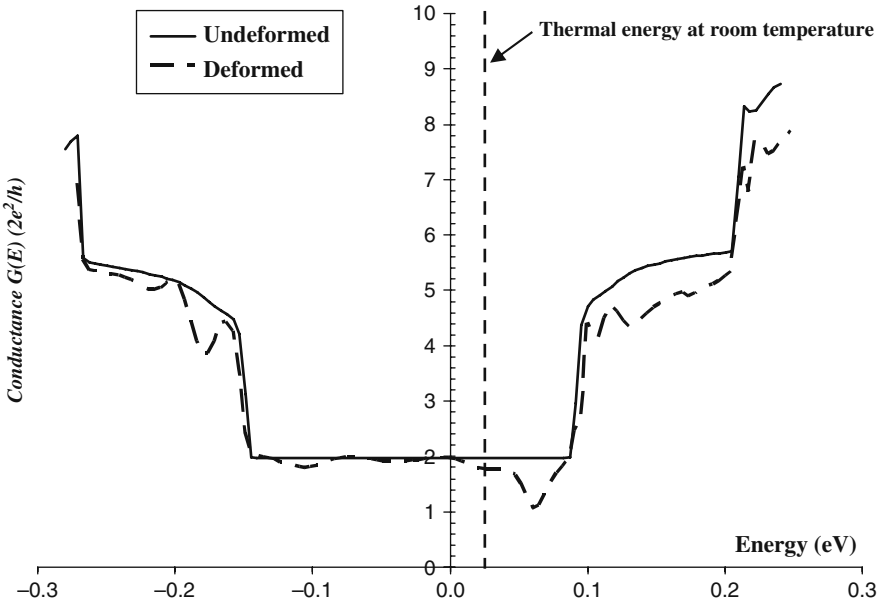


Figure 13-14. Conductance calculated using the mixed FE-TB approach for the model reproducing the MWNT used in the experiment, both in the undeformed and deformed configurations. The thermal energy at room temperature is also indicated

imposing a combination of axial compression and bending loading until the same configuration as that observed in the TEM image [66, 2b] was reached. Atomic coordinates, in both the undeformed and deformed configurations, were passed to the TB code for computing the transport properties of the MWNT. For a bias voltage of 1 V, the overall conductance of the undeformed tube obtained from the mixed FE-TB approach was the theoretical one, $G=2G_0$. At room temperature the thermal energy kT is about 0.025 eV.

Once the deformation level reached that shown in Figure 13-13, the TB code predicted $\sim$12% reduction in room-temperature conductance. These results are shown in Figure 13-14 as a plot of conductance versus energy, where the Fermi energy is taken as a reference and shifted to zero. Based on a computed conductance change similar to that reported in the experiments, the accuracy of the current predictions is considered to be rather good. We are unaware of any numerical simulations that have successfully replicated experimental results on electromechanical behavior of MWNTs.

13.3.7. Effect of the Outer Diameter on the Conductance of MWNTs Under Bending

With this validation of the computational approach, we now investigate the effect of MWNT outer diameter on the dependence of conductance on mechanical

deformation. In addition to the (70,70) MWNT described above, models of (50,50) and (30,30) armchair MWNTs were created and subjected to bending loading conditions. The new undeformed models resemble the (70,70) tube shown in Figure 13-13; the only differences consisted of removing the outer two walls of the (70,70) to produce the 6.78 nm outer-diameter (50,50) mesh, and the outer four walls for the 4.08 nm outer-diameter (30,30) mesh. The three nanotubes were bent to the same curvature by fixing the atoms of each wall near the two ends, as done in the previous simulation of the (70,70) MWNT, and then gradually rotating these rigid portions of the tube. A bending angle of $\theta = 1.0472$ radians (60°) is gradually applied, imposing global curvature $\kappa = \theta/L$; the final κ is 0.0214 (1/nm), since the deformable part of the model is 52 nm long. The deformed configurations are shown in Figure 13-15a–c.

For a bias voltage of 1 V, the computed overall conductance of all three undeformed tubes was the theoretical value, $G=2G_0$. Once the deformation levels reached those shown in Figure 13-15, the TB code predicted a reduction in conductance that varied over the considered energy range of –0.3 to 0.3 eV. Figure 13-16 shows the longitudinal sections of the (30,30), the (50,50), and (70,70) MWNTs at the bending angle of 60°. Figure 13-17 plots conductance versus energy for all three MWNTs, and the Fermi energy is again taken as a reference and shifted to zero. At the same curvature, the conductance of MWNTs is found to depend on the outer diameter. At room temperature, the smaller outer diameter exhibits

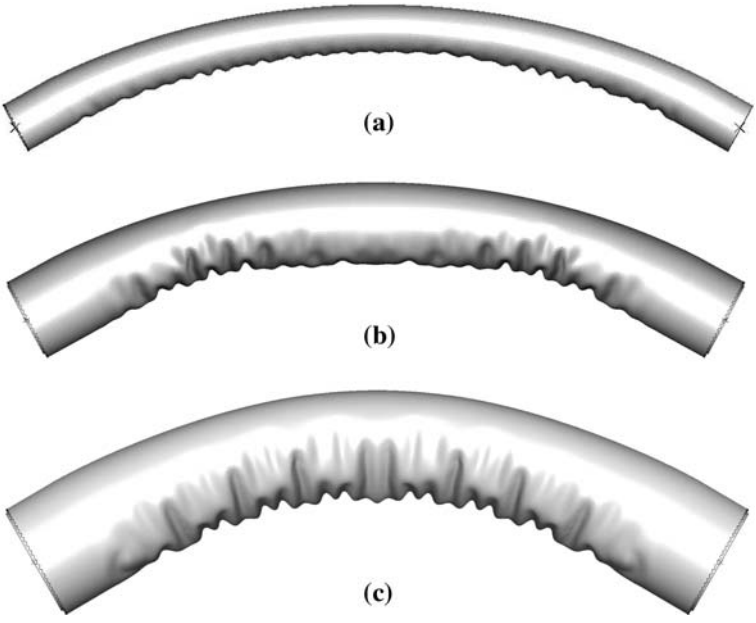


Figure 13-15. Views of the final deformed configurations of the FE models producing the same bending angle of 60° in (**a**) the (30,30), (**b**) the (50,50), and (**c**) the (70,70) MWNTs

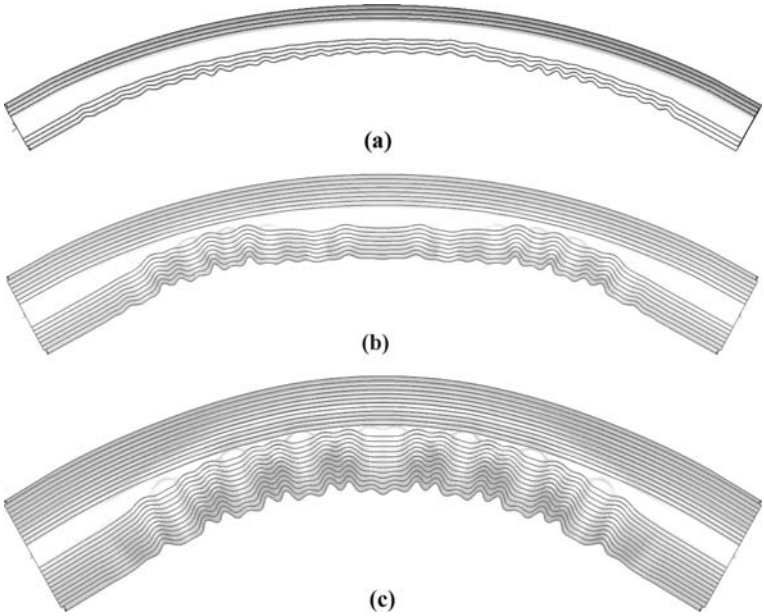


Figure 13-16. Sectional views of the final deformed configurations of the FE models producing the same bending angle of 60° in (**a**) the (30,30), (**b**) the (50,50), and (**c**) the (70,70) MWNTs

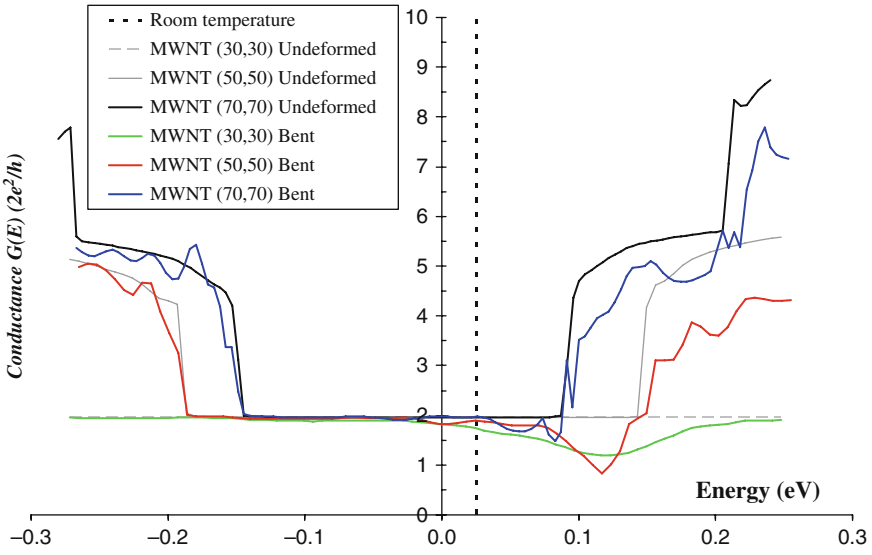


Figure 13-17. Conductance calculated using the mixed FE-TB approach for three armchair MWNT models of different outer diameters, both in the undeformed and deformed configurations

the greatest average drop in the conductance: the (30,30), the (50,50) and the (70,70) shows, respectively, a 13%, a 5.5% and a 0% reduction in the conductance. Additionally, the range in energy values where the conductance is reduced from the initial undeformed value, $G=2G_0$, is larger for the smaller diameter MWNTs.

Based on the outcomes of these simulations, we identified in the length of the circumference unaffected by the wrinkling one possible geometrical features of the deformed configuration able determine the dependence of conductivity on the MWNT diameter. In Table 13-1 the undistorted circumferential lengths of the three MWNTs are compared both in value and as percentage with respect to the total circumference of each tube. Even if the percentages are about the same, there is a big difference in the value of the lengths of the circumference unaffected by the wrinkling. In the (70,70) MWNT the undistorted circumferential length is 2.24 times longer than in the (30,30), and 1.366 longer than in the (50,50). In MWNTs, the nested van der Waals interactions of the increased number of inner walls radially stiffen the tube, thus limiting the circumferential extent of the kink. The number of relatively undistorted atoms outside the kinked portion of the circumference grows with the outer diameter. In the areas of the nantube walls where a kink is present, due to the distortion caused by the bending, centers of scattering are created that reduce the conductivity. At any kink, the local bonding structure is deformed and the electrons will lose some of the π character of their corresponding orbitals. As stated in [65], the curvature in the nanotube walls due to the kink leads to a loss of spatial overlap of the atomic p orbitals that contribute to conjugation and a shift in hybridization of the atoms from the sp^2 of graphite to something intermediate between sp^2 and sp^3. The net result of these orbital effects is an increase in energy locally and an introduction of partial radical character in the π-bonding electrons. Smaller nanotubes have a smaller undistorted circumferential length (smaller number of atoms in position to provide an effective transport pathway) than the larger diameter MWNTs, thus giving the corresponding larger influence on the loss in conductivity.

Table 13-1. Comparison of the undistorted circumferential lengths of the three MWNTs both in value and as percentage with respect to the total circumference of each tube

MWNT	Circumference of each tube (nm)	Length of the circumference unaffected by the wrinkling (nm)	Percentage of circumference unaffected by the wrinkling with respect to the total circumference
(30,30) MWNT	12.82	9.35	73
(50,50) MWNT	21.3	15.33	72
(70,70) MWNT	29.82	20.95	70

13.3.8. Effect of the Outer Diameter on the Conductance of MWNTs Under Stretching

Here we investigate the one of the possible effect of MWNT outer diameter on the dependence of conductance on mechanical deformation. Models of (70,70), (50,50) and (30,30) armchair MWNTs were created and subjected to traction loading conditions. All the models are 60 nm long, while the diameter of the outermost tube of the (70,70), (50,50) and (30,30) armchair MWNTs are, respectively, 9.492, 6.78 and 4.08 nm. The three nanotubes were stretched out by fixing the atoms of each wall near the two ends, then displacing these 4 nm long rigid portions of the tubes. An extension of 2.55 nm was gradually applied, imposing global axial deformation of 0.0472, since the deformable part of the model is 52 nm long. The deformed configurations are shown in Figures 13-18 and 13-19. Clearly these are not pure tension boundary conditions, but an attempt to reproduce a possible real working

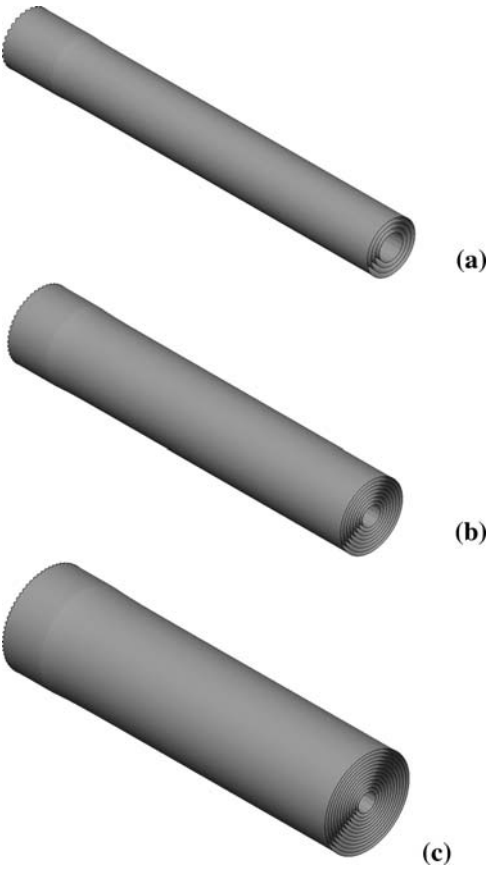


Figure 13-18. Views of the final 3D deformed configurations of the FE models in (**a**) the (30,30), (**b**) the (50,50), and (**c**) the (70,70) MWNTs. The figure shows only half of the carbon nanotube models

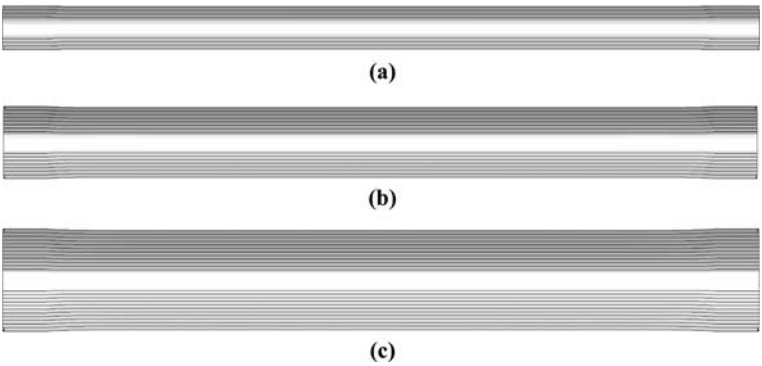


Figure 13-19. Sectional views of the final deformed configurations of the FE models in (**a**) the (30,30), (**b**) the (50,50), and (**c**) the (70,70) MWNTs

condition for MWNTs bases electronic devices where the ends of the nanotubes are fixed inside metallic leads and cannot deform. For a bias voltage of 1 V, the computed overall conductance of all three undeformed tubes was the theoretical value, $G=2G_0$. Once the deformation levels reached those shown in Figures 13-18 and 13-19, the TB code predicted a reduction in conductance that varied over the considered energy range of –0.1 to 0.3 eV. The Fermi energy is again taken as a reference and shifted to zero. At the same extension, the conductance of MWNTs is found unchanged in the energy range where for of all three undeformed tubes it was the theoretical value, $G=2G_0$. While when these energy levels are exceeded there is a significant reduction in the conductance of the deformed tubes. Since the energy level at which conductance have the theoretical value of $2G_0$ is lower for bigger armchair MWNTs, we can conclude that for this particular load case armchair carbon nanotubes with smaller outer diameter maintain their conductance unchanged up to higher energy levels (Figure 13-20).

13.3.9. Effect of Current Saturation – Non-Linear I-V Response

Theoretical and experimental results have proved that conductance in carbon nanotubes, both SWNTs and MWNTs, drops dramatically as the bias voltage is increased due to scattering of electrons [e.g., 12, 67–71]. As discussed in, [68] the current saturation derives from the band structure. In metallic CNTs current is carried by two propagating 1D sub-bands. When scattering is not present, the difference between the chemical potentials of the right and left moving states will be given by the applied voltage. When the bias voltage is low the I-V response is Ohmic, but for high voltage level when eV exceeds the Fermi energy of the 1D sub-bands, the left moving states will be completely depleted and the current will saturate. In experiments the current saturation will start at a much lower voltage with respect to theoretical predictions.

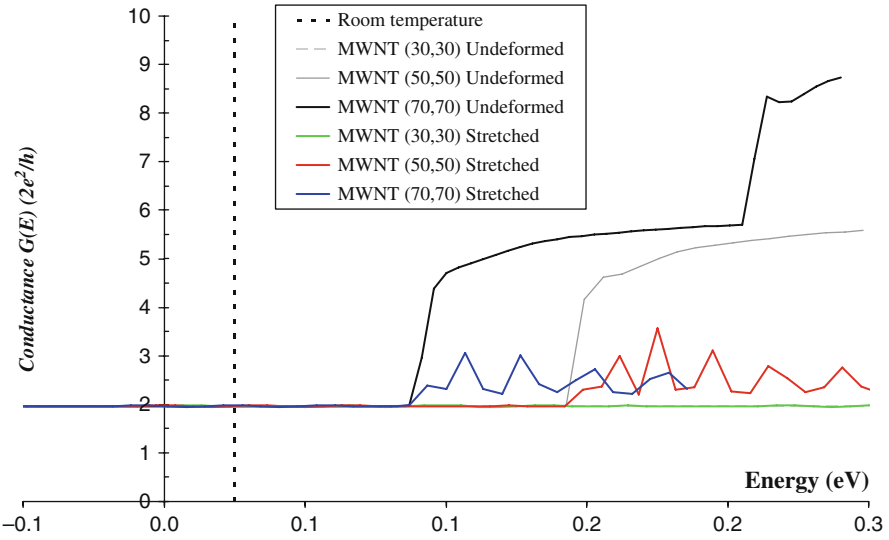


Figure 13-20. Conductance calculated using the mixed FE-TB approach for three armchair MWNT models of different outer diameters, both in the undeformed and deformed configurations

Yao et al. [68] using low-resistance electrical contacts have measured the high-field transport properties of metallic SWNTs. Their experimental results [Figure 2, 40] show a linear I-V response for bias volage under 1 V, at higher voltage current saturation leads to a quick reduction in conductance that determines a plateau in the I-V curve. Yao et al. [68] also discussed if in the experiments the contact resistance maybe the primary cause of the drop in the conductance. They concluded that the current saturation cannot derive from an amplified contact resistance as the voltage grow because the contacts should then behave like high-resistance tunneling contacts, and in the I-V associated with tunneling should appear features that are not present in the measured I-V.

Similar results have been found for MWNTs [e.g., [67, 69–71]. For example, Collins et al. [69] investigated high energy transport in MWNTs with a test structures are produced by depositing arcgrown MWNTs onto SiO_2 substrates prepatterned with Au/Ti electrodes. Their I-V measurements to the high current [Figure 2, 69] show an extremely nonlinear regime. Above the low bias, linear regime, MWNT I-V curves have a point of inflection above which dI/dV decreases towards a sample-dependent saturation.

Using the FE-TB code we have predicted the change in conductance of the 4-walled MWNT whose outermost wall is made of a (30,30) armchair CNT, as described in Chapter 13.3.7. The conductance of the MWNT at various bias voltage level, shown in Figure 13-21, has been calculated in the undeformed configuration and in the deformed bent configuration of Chapter 13.3.7. Since our calculations are theoretical, our results are expected to leave the linearity later that the experimental results.

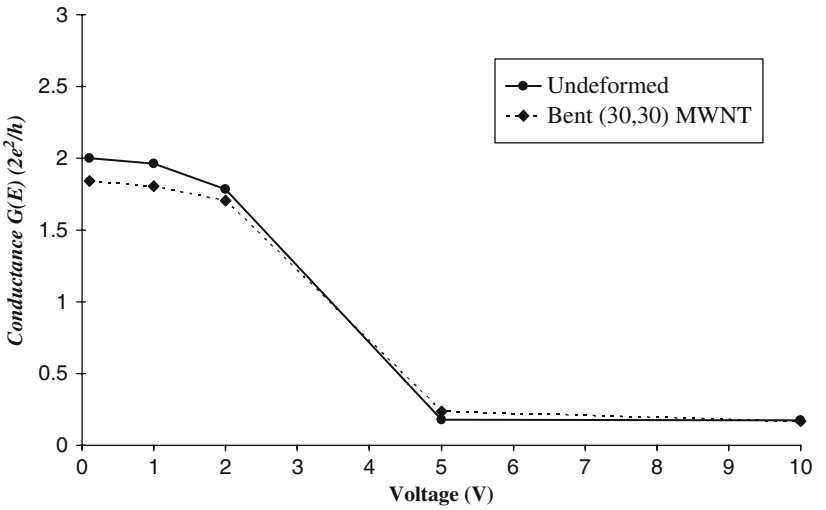


Figure 13-21. Conductance calculated using the mixed FE-TB approach for (30,30) armchair MWNT model at various bias voltage level in the deformed bent configuration at the Fermi level

13.4. CONCLUSIONS

In summary, a computationally-effective mixed finite element-tight-binding approach has been developed which can simulate the electromechanical behavior of SWNTs and MWNTs of the dimensions used in nano-electronic devices. The TB code is carefully designed to realize orders-of-magnitude reduction in computational time in calculating deformation-induced changes in electrical transport properties of the nanotubes. Finally, the effect of the MWNT diameter on the conductance after mechanical deformation was investigated.

REFERENCES

1. Poncharal P, Wang ZL, Ugarte D, de Heer WA (1999) Science 283: 1513–1516
2. Wang ZL, Poncharal P, de Heer WA (2000) J. Phys. Chem. Solids 61: 1025–1030
3. Yu MF, Lourie O, Dyer MJ, Moloni K, Kelly TF, Ruoff RS (2000) Science 287: 637–640
4. Wong EW, Sheehan PE, Lieber CM (1997) Science 277: 1971–1973
5. Qi HJ, Teo KBK, Lau K, Boyce MC, Milne WI, Robertson J, Gleason KK (2003) J. Mech. Phys. Solids 51: 2213–2237
6. Robertson DH, Brenner DW, Mintmire JW (1992) Phys. Rev. B 45: 12592–12595
7. Zhou X, Zhou JJ, Ou-Yang ZC (2000) Phys. Rev. B 62: 13692–13696
8. Sanchez-Portal D, Artacho E, Soler J, Rubio A, Ordejon P (1999) Phys. Rev. B 59: 12678–12688
9. Yakobson B, Brabec C, Bernholc J (1996) Phys. Rev. Lett. 76: 2511–2514
10. Tu Z, Ou-Yang Z (2002) Phys. Rev. B 65: 233407
11. Kundin KN, Scuseria GE, Yakobson BI (2001) Phys. Rev. B 64: 235406
12. Kelly B (1981) Physics of Graphite. Applied Science Publishers, London
13. Pantano A, Boyce MC, Parks DM (2004) J. Mech. Phys. Solids 52: 789–821

14. Pantano A, Parks DM, Boyce MC (2003) Phys. Rev. Lett. 91: 145504
15. Pantano A, Boyce MC, Parks DM (2004) J. Eng. Mater. Technol. 126: 279–284
16. Bower C, Rosen R, Jin L, Han J, Zhou O (1999) Appl. Phys. Lett. 74: 3317–3319
17. Lourie O, Cox DM, Wagner HD (1998) Phys. Rev. Lett. 81: 1638–1641
18. Cumings J, Zettl A (2000) Science 289: 602–604
19. Saito R, Dresselhaus G, Dresselhaus MS (1998) Physical Properties of Carbon Nanotubes. Imperial College Press, London; Dresselhaus MS, Dresselhaus G, Avouris P (2001) Carbon Nanotubes. Springer, Heidelberg
20. Bernholc J, Brenner D, Buongiorno Nardelli M, Meunier V, Roland C (2002) Annu. Rev. Mater. Res. 32: 347
21. Kong J, Yenilmez E, Tombler TW, Kim W, Dai H, Laughlin RB, Liu L, Jayanthi CS, Wu SY (2001) Phys. Rev. Lett. 87: 106801
22. Yu MF, Lourie O, Dyer MJ, Moloni K, Kelly TF, Ruoff RS (2000) Science 287: 637
23. Tekleab D, Carroll DL, Samsonidze GG, Yakobson BI (2001) Phys. Rev. B 64: 035419
24. Frank S, Poncharal P, Wang ZL, de Heer WA (1998) Science 280: 1744
25. Gupta R, Smallcup RE, in het Panhuis M (2005) Nanotechnology 16: 1707
26. Schönenberger C, Bachtold A, Strunk C, Salvetat JP, Forró L (1999) Appl. Phys. A 69: 283
27. Batchtold A, Strunk C, Salvetat JP, Bonard JM, Salvetat JP, Nussbaumer T, Schönenberger C (1999) Nature 397: 673
28. Bourlon B, Miko C, Forro' L, Glattli DC, Bachtold A (2004) Phys. Rev. Lett. 93: 176806
29. Roche S, Triozon F, Rubio A, Mayou D (2001) Phys. Lett. A 285: 94
30. Collins PG, Avouris Ph (2002) Appl. Phys. A 74: 329–332
31. Paulson S, Falvo MR, Snider N, Helser A, Hudson T, Seeger A, Taylor RM, Superfine R, Washburna S (1999) Appl. Phys. Lett. 75: 2936
32. Minot ED, Yaish Y, Sazonova V, Park J, Brink M, McEuen PL (2003) Phys. Rev. Lett. 90: 156401
33. Liu B, Jiang H, Johnson HT, Huang Y (2004) J. Mech. Phys. Solids 52: 1
34. Maiti A, Svizhenko A, Anantram MP (2002) Phys. Rev. Lett. 88: 126805
35. Lu J, Wu J, Duan W, Liu F, Zhu B, Gu B (2003) Phys. Rev. Lett. 90: 156601
36. Farajian AA, Yakobson BI, Mizuseki H, Kawazoe Y (2003) Phys. Rev. B 67: 205423
37. Rochefort A, Avouris P, Lesage F, Salahub DR (1999) Phys. Rev. B 60: 13824
39. Yang L and Han J (2000) Phys. Rev. Lett. 85: 154
40. Tunney MA, Cooper NR (2006) Phys. Rev. B 74: 075406
41. Sanvito S, Kwon Y, Tománek D, Lambert CJ (2000) Phys. Rev. Lett. 84: 1974
42. Delaney P, Di Ventra M, Pantelides ST (1999) Appl. Phys. Lett. 75: 3787
43. Kwon Y, Tománek D (1998) Phys. Rev. B 58: R16001
44. Choi HJ, Ihm J, Yoon Y, Louie SG (1999) Phys. Rev. B 58: R14009
45. Baughman RH, Zakhidov AA, de Heer WA (2002) Science 297: 787
46. Bernholc J, Brabec C, Nardelli M, Maiti A, Roland C, Yakobson B (1998) Appl. Phys. A Mater. Sci. Process. 67: 39–46
47. Belytschko T, Xiao S, Schatz G, Ruoff R (2002) Phys. Rev. B 65: 235430
48. Yakobson B, Avouris P (2001) Springer, Heidelberg, pp. 287, M.S. Dresselhaus et al., eds.
49. Girifalco LA, Hodak M, Lee RS (2000) Phys. Rev. B 62: 13104–13110
50. Girifalco LA, Lad RA (1956) J. Chem. Phys. 25: 693–697
51. Zhao YX, Spain IL (1989) Phys. Rev. B 40: 993–997
52. Ruoff RS, Ruoff AL (1991) Appl. Phys. Lett. 59: 1553–1555
53. Qian D, Liu WK, Ruoff RS (2001) J. Phys. Chem. B 105: 10753–10758
54. Hanfland M, Beister H, Syassen K (1989) Phys. Rev. B 39: 12598–12603
55. Boettger J (1997) Phys. Rev. B 55: 11202–11211
56. Ru CQ (2000) J. Appl. Phys. 87: 7227–7231

57. Xu CH, Wang CT, Chan CT, Ho KM (1992) J Phys Condens Matter 4: 6047
58. Charlier J-C, Lambin Ph, Ebbesen TW (1997) Phys. Rev. B 54: R8377
59. Porezag D, Frauenheim Th, Kohler Th, Seifert G, Kaschner R (1995) Phys. Rev. B 51: 12947
60. Meunier V, Buongiorno Nardelli M, Roland C, Bernholc J (2001) Phys. Rev. B 64: 195419
61. Hertel T, Walkup R, Avouris P (1998) Physical Review B 58: 13870–13873
62. Timoshenko S (1936) Theory of Elastic Stability. McGraw–Hill, New York
63. Falvo M, Clary G, Taylor R, Chi V, Brooks F, Washburn S, Superfine R (1997) Nature 389: 582–584
64. Maiti A (2000) Chemical Phys. Lett. 331: 21–25
65. Ebbesen TW, Takada T (1995) Carbon 33: 973–978
66. Kuzumaki T, Mitsuda Y (2004) Appl. Phys. Lett. 85: 1250–1252
67. Charlier JC, Blase X, Roche S (2007) Rev Mod Phys 79: 677
68. Yao Z, Kane CL, Dekker C (2000) Phys. Rev. Lett. 84: 2941
69. Collins PG, Arnold M, Hersam M, Martel R, Avouris P (2001) Phys. Rev. Lett. 86: 3128
70. Collins PG, Arnold M, Avouris P (2001) Science 292: 706
71. Yao Z, Postma H, Balents L, Dekker C (1999) Nature 402: 273

CHAPTER 14

MULTISCALE MODELING OF CARBON NANOTUBES

YUZHOU SUN[1,2] AND K.M. LIEW[1]

[1]*Department of Building and Construction, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, e-mail: kmliew@cityu.edu.hk*

[2]*Department of Civil Engineering and Architecture, Zhongyuan University of Technology, Zhengzhou 450007, China*

Correspondance to K. M. Liew, kmliew@cityu.edu.hk

Abstract: Theoretical modeling and simulation play an important role in understanding the subtle and complex behavior of carbon nanotubes (CNTs). Atomic simulations can capture the microscale mechanism of nanostructures, but they are limited to very small systems due to their huge computational cost. Continuum simulations can extend to large enough systems, however they can't reflect microscale physical laws of nanostructures. Multiscale modeling that couples atomic simulations and continuum modeling is emerging as a feasible and efficient approach for large-size nanostructures. This chapter aims to systematically illustrate the three components of multiscale modeling of CNTs: atomic simulation, continuum modeling approach, and multiscale coupling scheme. The chapter first reviews several multiscale coupling schemes, and then introduces an atomic modeling approach and a higher-order continuum model. The mesh-free method is employed to implement the continuum discretization, and multiscale analysis is achieved by appropriately coupling the mesh-free continuum framework and the atomic simulation. Computations are carried out for CNTs using atomic simulations, continuum modeling and multiscale analysis, respectively, and the efficiency of multiscale modeling is discussed

Keywords: Carbon nanotubes, Multiscale modelling, Brenner potential, Atomic simulation, Continuum method, Cauchy-Born rule, Higher-order continuum, Buckling

14.1. INTRODUCTION

Carbon nanotubes (CNTs) were first discovered by Iijima [1] of the NEC Corporation in Japan in 1991. Due to their unique nanostructure, CNTs possess remarkable physical, electrical, chemical, and mechanical properties, and have a wide range of potential applications in nano-electronics, quantum wire interconnects, field emission panel display, composites, chemical sensors, biosensors, detectors, etc. In addition to a large amount of experimental work, theoretical modeling plays an important role in capturing and understanding the delicate

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 367–388.
DOI 10.1007/978-1-4020-9785-0_14,

behavior of nanostructures. Modeling approaches can be generally classified into two categories: atomic simulation such as molecular dynamics [2–4] (MD) and continuum method [5–13]. Atomic simulations can capture the microscale mechanism of nanostructures and yield results that are in many cases explicit in nature. However, atom-based methods consume a large amount of computational resources, and thus computation is limited to a very small size. This huge computational cost largely restricts the use of atomic-based methods, which has led to the emergence of continuum method, which makes use of the continuum mechanics theory to study CNTs. Continuum-based methods are much faster than atomic simulation in the analysis of systems of engineering interest, which makes them attractive. Continuum simulation can largely reduce the degrees of freedom in problems, and the theoretical and numerical analysis of large-scale structures thus become possible. In addition, continuum simulation can also display certain CNT properties that cannot be found with atomic simulations. Some equivalent continuum models, for example, the beam model [5], the shell model [6, 7], and the truss/frame model [8, 9], have been developed, and have proved to be very efficient from the computational point of view. However, continuum simulations can't reflect the microscale physical laws of nanostructures, and are not adequate for the analysis of CNTs.

The limitations of atomic simulations as well as continuum simulations have stimulated extensive research into multiscale methods that bridge atomic simulation and continuum descriptions. Multiscale method takes advantage of both approaches and can overcome the length and time scale limits in an efficient manner, and it is emerging as a feasible and efficient approach for large-size problems. The basic idea of the multiscale method is to use atomic simulation for the localized region in which the discrete motion of atoms is important and to use the continuum method for the remaining regions in which the deformation is considered to be homogeneous and smooth. A key issue associated with the multiscale method is the way to bridge the two different scales smoothly. In the early multiscale study of CNTs, almost all researches are focused on exploring the appropriate model to smoothly couple two scales. Currently, several efficient coupling approaches, example for the quasi-continuum method [14, 15], the bridging domain method [16, 17], and the bridging scale method [18, 19], have been proposed.

From the above discussion, it can be seen that there are three components in a multiscale modeling: an atomic method, a continuum modeling approach, and an appropriate coupling scheme. These three components will respectively be introduced in this chapter. Several multiscale coupling approaches are first reviewed in the Section 14.2. Although MD is a commonly popular atomic simulation, we introduce a faster atomic modeling approach that has the same solving scheme as the subsequent mesh-free continuum simulation. In the previous multiscale modeling of CNTs, the researchers always ignored the importance of the rationality of the continuum models. This work introduces a reasonable and accurate continuum model for CNTs. The multiscale modeling is carried out by coupling the mesh-free computational framework and the atomic simulation. The present work aims to present a comprehensive understanding for the multiscale modeling of CNTs.

14.2. MULTISCALE COUPLING APPROACHES

Depending on the methods of coupling the multiple scales, multiscale simulation of CNTs can be either concurrent or hierarchical. In the concurrent simulation, two scale methods are coupled within one unified numerical code in which the bridging technique provides the link among them. In the hierarchical approach, simulations are carried out at separate scales, which provide critical insights for improved modeling in the next larger scale. To achieve the concurrent simulation, domain decomposition is generally used. Atomic simulation is implemented in the localized region where the discrete motion of atoms is important, and the continuum method is used for the remaining regions in which the deformation is homogeneous and smooth. Such a multiscale scheme can capture the quantities that vary quickly in the critical atomic region while significantly reducing the computational cost by treating the surrounding material in an averaged sense. A great challenge in multiscale simulations is the treatment of the linking area between two scales. Several techniques have been developed to couple the continuum and atomic methods. Here, three popular approaches, the quasi-continuum method [14, 15], the bridging domain method [16, 17], and the bridging scale method [18, 19], are introduced.

14.2.1. Quasi-Continuum Method

This method was proposed by Tadmor et al. [14, 15] in modeling 2-D crystallites, and it has been used to study a variety of fundamental aspects of deformation in crystalline solids, including fracture, grain boundary structure and deformation, nano-indentation, and 3-D dislocation junctions.

The quasi-continuum method introduced the concept of representative atoms. Shown in Figure 14-1a is a crystalline solid that involves a Lomer dislocation [20]: the filled circles denote the chosen representative atoms. Around the core of the dislocation, the deformation is drastic, the microcosmic change of the atomic structure needs to be described and traced, and all atoms are chosen as the representative atoms. Far away from the core of the dislocation, the field is slowly varying and the deformation is homogeneous, and certain sparse atoms are chosen as the representative atoms. In the former region, the atoms are called non-local atoms, and their interactions with the neighboring atoms are reflected. The atoms in the latter region are called local atoms, and a continuum discretization scheme can be established based on them.

The total potential energy of the quasi-continuum model can be obtained as

$$E_{\mathrm{QC}} = \sum_{\text{nonlocal}} V_i + \sum_{\text{local}} w_\mu V_\mu \cdot \tag{14-1}$$

The first and second terms in Eq. (14-1) denote the energies of the non-local and local atoms, respectively. w_μ is the weight function and refers to the number of atoms that this representative atom can represent. In practice, the local region can

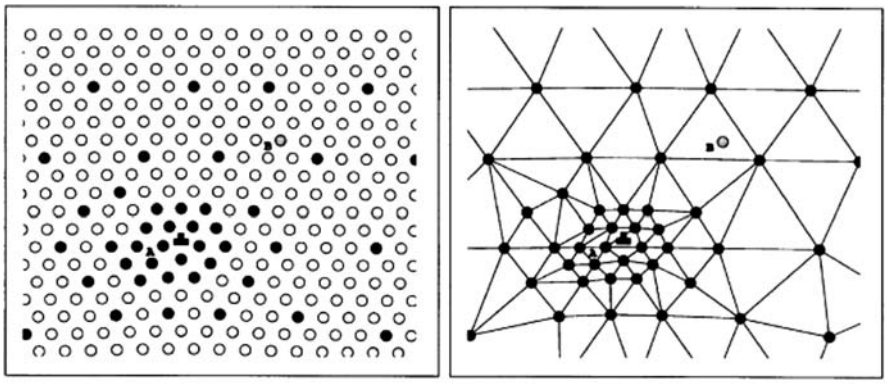


Figure 14-1. (**a**) Around the core of a Lomer dislocation, all atoms are chosen as the representative atoms (*filled circles*), whereas sparse atoms are chosen the far field. (**b**) The constructed FEM based on the chosen representative atoms [20]

be discretized as a series of elements by using the representative atoms as nodes (see Figure 14-1b).

In the quasi-continuum method, a ghost force often arises in the transition zone between the local and non-local regions, even when the crystal is undeformed. To avoid the effect of this ghost force, the energy is augmented by a term associated with the work done by the ghost force:

$$\hat{E}_{\mathrm{QC}} = E_{\mathrm{QC}} - \sum_{g} f^{g} u_{g} \cdot \tag{14-2}$$

The ghost force f^{g} is calculated each time the status of the representative atoms is updated and is held constant until an update is required due to the evolving state of the deformation [20].

14.2.2. Bridging Domain Method

The bridging domain method was proposed by Belytschko and Xiao [16, 17], was used to simulate the fracture of graphite sheets [17], and was recently extended to simulate the fracture of large-diameter CNTs [17, 21]. In this technique, the entire domain is decomposed into three types of regions (see Figure 14-2): an atomic region Ω^{A} in which the atomic movement is described, a continuum region $\Omega^{\mathrm{C}} = \Omega^{\mathrm{CL}} + \Omega^{\mathrm{CR}}$ in which the lattice undergoes a smooth deformation and the continuum simulation is used, and an overlapping region $\Omega^{\mathrm{O}} = \Omega^{\mathrm{A}} \cap \Omega^{\mathrm{C}}$ in which the atomic and continuum models overlap. In the atomic and continuum regions, the energy is computed using the atomic and continuum methods, respectively. In the overlapping region, the potential energy is expressed as a linear combination

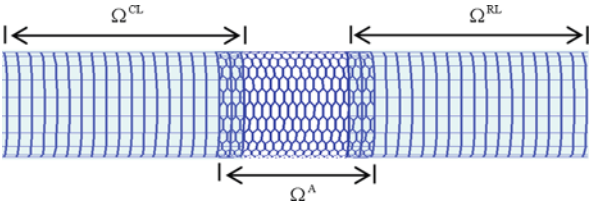


Figure 14-2. The entire domain is decomposed into a continuum simulation region $\Omega^{C} = \Omega^{CL} + \Omega^{CR}$ and a full atomic region Ω^{A} with an overlapping region $\Omega^{O} = \Omega^{A} \cap \Omega^{C}$

of the continuum and atomic energies. This ensures smooth bridging between the continuum and atomic deformation fields.

14.2.3. Bridging Scale Method

The bridging scale method was developed by the research group of Liu [18, 19]. In this method, the total displacement field is decomposed into two different components, as follows.

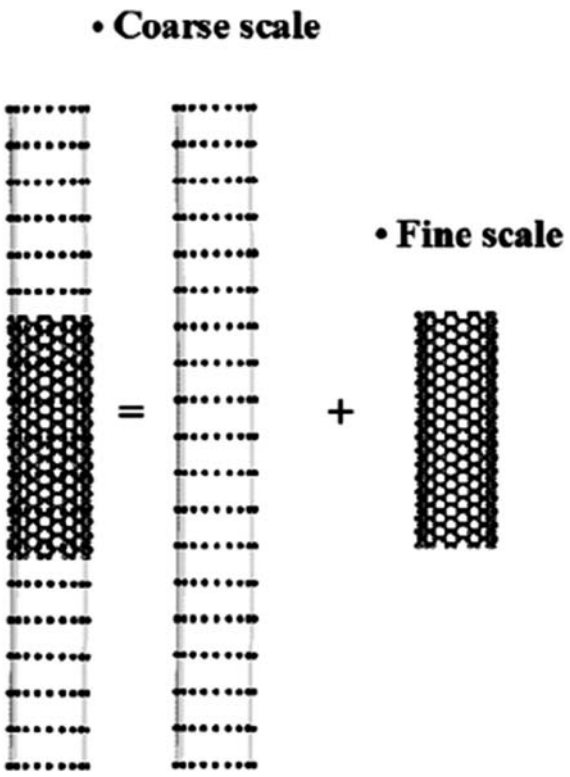


Figure 14-3. Schematic illustration of the multiscale discretization scheme in the bridging scale method [18]

$$\mathbf{u}(\mathbf{x}) = \overline{u}(\mathbf{x}) + \mathbf{u'}(\mathbf{x}), \tag{14-3}$$

where $\overline{u}(\mathbf{x})$ is the coarse scale component and can be established in a continuum discretization frame such as the finite element or mesh-free method. $\mathbf{u}'(\mathbf{x})$ is the fine scale component that must be determined by the atom-based method. Far away from the localized domain, $\mathbf{u}'(\mathbf{x})$ becomes insignificant, or, mathematically, $\mathbf{u}'(\mathbf{x})$ has a vanishing projection onto the coarse scale basis function. Therefore, atomic simulation is only needed in the localized regions. Figure 14-3 shows the computational scheme for the bridging scale method. There are two different computational domains. The first is the coupled domain in which the atomic structure co-exists with the continuum discretization. The second domain is the coarse scale domain in which only the continuum discretization exists. The coarse and fine scale solutions are coupled. In the computation, the coarse scale solution is solved in the entire domain, whereas the fine scale solution is only solved in the localized region. Bridging between the coarse and fine scales is realized by transparently exchanging information between the coarse and fine scale regions.

14.3. BRENNER POTENTIAL

In the theoretical and numerical modeling of CNTs, the empirical potential is always used to describe the interaction between carbons. The Brenner potential [22, 23] was first proposed for hydrocarbons, and has been widely used in the study of CNTs. In the Brenner potential, the system energy means the sum of the energy on each bond, and the energy of each bond is composed of a repulsive and attractive pair. Following Tersoff [22] and Brenner [23], the expression for the bonding energy between atoms I and J is

$$V_B(r_{IJ}) = V_R(r_{IJ}) - B_{IJ}V_A(r_{IJ}), \tag{14-4}$$

where r_{IJ} is the distance between atoms I and J, and V_R and V_A are the repulsive and attractive pairs of the energy terms given by

$$V_R(r_{IJ}) = \frac{D^{(e)}}{S-1} e^{-\sqrt{2S}\beta(r_{IJ}-R^{(e)})} f_c(r_{\mathrm{IJ}}), \tag{14-5}$$

$$V_A(r_{IJ}) = \frac{D^{(e)}S}{S-1} e^{-\sqrt{2/S}\beta(r_{IJ}-R^{(e)})} f_c(r_{IJ}). \tag{14-6}$$

f_c is a smooth cut-off function that limits the range of the potential, and is given by

$$f_c(r) = \begin{cases} 1 & r < R^{(1)} \\ \frac{1}{2}\left\{1 + \cos\left[\frac{\pi(r-R^{(1)})}{R^{(2)}-R^{(1)}}\right]\right\} & R^{(1)} < r < R^{(2)} \\ 0 & r > R^{(2)} \end{cases} \tag{14-7}$$

where $R^{(1)}$ and $R^{(2)}$ are the effective ranges of the cut-off function.

The parameter B_{IJ} in Eq. (14-4) represents multi-body coupling between the bond $I-J$ and the local environment of atom I, and is given by

$$B_{IJ} = \left[1 + \sum_{K \neq I,J} G(\theta_{IJK}) f_c(r_{IK}) \right]^{-\delta}, \tag{14-8}$$

where θ_{IJK} is the angle between the bonds $I-J$ and $I-K$, and the angle function G is given by

$$G(\theta_{IJK}) = a_0 \left[1 + \frac{c_0^2}{d_0^2} - \frac{c_0^2}{d_0^2 + (1 + \cos(\theta_{IJK}))^2} \right]. \tag{14-9}$$

For atoms I and J with different local environments, Brenner [23] suggested replacing B_{IJ} in Eq. (14-8) with

$$\bar{B}_{IJ} = \frac{1}{2}(B_{IJ} + B_{JI}). \tag{14-10}$$

The parameters $D^{(e)}, S, \beta, R^{(e)}, \delta, R^{(1)}, R^{(2)}, a_0, c_0, d_0$ in the above equations can be determined by fitting with the known physical properties of carbon. Brenner gave two sets of parameters as follows [23]:

(I) $D^{(e)} = 6.325$ eV, $S = 1.29$, $\beta = 15\ \text{nm}^{-1}$, $R^{(e)} = 0.1315$ nm, $R^{(1)} = 0.17$ nm, $R^{(2)} = 0.2$ nm, $\delta = 0.80469$, $a_0 = 0.011304$, $c_0 = 19$, $d_0 = 2.5$; (14-11)

and

(II) $D^{(e)} = 6.000$ eV, $S = 1.22$, $\beta = 21\ \text{nm}^{-1}$, $R^{(e)} = 0.139$ nm, $R^{(1)} = 0.17$ nm, $R^{(2)} = 0.2$ nm, $\delta = 0.50000$, $a_0 = 0.00020813$, $c_0 = 330$, $d_0 = 3.5$. (14-12)

The Brenner potential has recently been revised [24], and the new version is generally called as the second-generation Brenner potential. In comparison with the Brenner potential, the second-generation potential includes both improved analytic functions for the interatomic interactions and an extended fitting database, which results in a significantly better description of the bond length, energies, and force constants for hydrocarbon molecules and the elastic properties, interstitial defect energies, and surface energy for diamonds.

The second-generation potential is also expressed as Eq. (14-4), and V^R and V^A are given as

$$V^R(r_{IJ}) = f^c(r_{IJ}) \left(1 + \frac{Q}{r_{IJ}} \right) A e^{-\alpha r_{IJ}}, \tag{14-13}$$

$$V^A(r_{IJ}) = f^c(r_{IJ}) \sum_{n=1,3} B_n e^{-\beta_n r_{IJ}}, \tag{14-14}$$

where Q is the effective charge in the screened coulomb potential, and $f^c(r_{IJ})$ is a cutoff function that is similar to $f_c(r_{IJ})$. The detail on the terms B_{IJ}^{DH}, A, α, B_n and β_n can be found in Ref. [24].

In simulations of the failure of CNTs that use the second-generation Brenner potential, researchers [19, 21] have found that the fracture stresses are several times larger than those of quantum mechanical results. This is due to the functional form of the cutoff function in the potential, which artificially raises the bond force for the distances between 1.7 and 2.0. To avoid nonphysical failure mechanisms, it is suggested that this cutoff function be removed in the fracture analysis, but included in the *C-C* interactions only for those atom pairs that are less than 2.0Å apart in the initial and undeformed configurations [19, 21]. With this modification, the potential is no longer capable of handling bond formation, but it can give reasonable results for the fracture. In the present research, the Brenner potential with the second set of parameters is mainly used, except the fracture simulation of CNTs (Section 14.7.2) in which we use the modified second-generation Brenner potential.

14.4. AN ATOMIC SIMULATION METHOD

The multiscale method requires an appropriate atomic simulation approach. The commonly popular atomic method is molecular dynamic. Here, we introduce a different approach that has the same solution framework as the later mesh-free method. This atomic simulation method will be employed in our multiscale simulation.

Let us consider a system that contains *NA* carbon atoms. The energy that is stored in the atomic bonds is a function of the positions of all atoms, as follows.

$$U_{nb} = U_{nb}(\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_N) = \sum_{I<J}^{NA} V_B(\mathbf{q}_J - \mathbf{q}_I), \tag{14-15}$$

where $\mathbf{q}_I$ is the position of atom I, and $V_B(\mathbf{q}_J - \mathbf{q}_I)$ is the Brenner potential.

The total potential energy of the atomic system is

$$E_{\mathrm{A}}(\mathbf{q}) = U_{nb}(\mathbf{q}) - \sum_{I=1}^{NA} \bar{f}_I \cdot \mathbf{q}_I, \tag{14-16}$$

where $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_{NA})^{\mathrm{T}}$, and $\bar{f}_I$ is the external force (if any) that is exerted on atom I. The state of the stable configuration corresponds to

$$\frac{\partial E_A(\mathbf{q})}{\partial \mathbf{q}} = 0. \tag{14-17}$$

The Taylor expansion of $E_{\mathrm{A}}(\mathbf{q})$ around an initial guess of $\mathbf{q}^0 = (\mathbf{q}_1^0, \mathbf{q}_2^0, \cdots, \mathbf{q}_{NA}^0)^{\mathrm{T}}$ for the equilibrium state gives

$$E_{\mathrm{A}}(\mathbf{q}) \approx E_{\mathrm{A}}(\mathbf{q}^0) + \left.\frac{\partial E_{\mathrm{A}}(\mathbf{q})}{\partial \mathbf{q}}\right|_{\mathbf{q}=\mathbf{q}^0} \cdot (\mathbf{q} - \mathbf{q}^0) + \frac{1}{2}(\mathbf{q} - \mathbf{q}^0)^{\mathrm{T}} \cdot \left.\frac{\partial E_{\mathrm{A}}(\mathbf{q})}{\partial \mathbf{q}\partial \mathbf{q}}\right|_{\mathbf{q}=\mathbf{q}^0} \cdot (\mathbf{q} - \mathbf{q}^0). \tag{14-18}$$

Its substitution into Eq. (14-17) yields the following governing equation for the displacement $\Delta\mathbf{q} = (\mathbf{q} - \mathbf{q}^0)$.

$$\mathbf{K}_{\mathrm{A}} \Delta\mathbf{q} = \mathbf{f}, \tag{14-19}$$

where the stiffness matrix is

$$\mathbf{K}_{\mathrm{A}} = \frac{\partial^2 E_{\mathrm{A}}}{\partial \mathbf{q}\partial \mathbf{q}} = \frac{\partial^2 U_{nb}}{\partial \mathbf{q}\partial \mathbf{q}}, \tag{14-20}$$

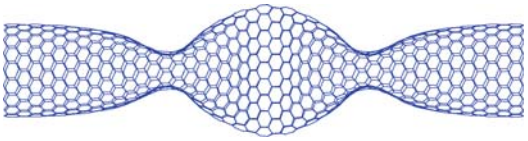
and the non-equilibrium force vector is

$$\mathbf{f} = -\frac{\partial E_{\mathrm{A}}}{\partial \mathbf{q}} = \bar{f} - \frac{\partial U_{nb}}{\partial \mathbf{q}}, \tag{14-21}$$

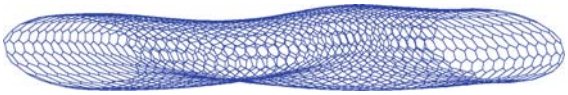
where $\bar{f} = (\bar{f}_1, \bar{f}_2, \cdots, \bar{f}_{NA})^{\mathrm{T}}$ is the external force vector. The equilibrium state is obtained by iteratively solving Eq. (14-19) until $\mathbf{f}$ reaches zero.

Liu et al. [25, 26] recently introduced a concept of the atomic-scale finite element to assemble the global stiffness matrix and force vector, and they employed finite element software to solve large-scale problems. In the present computation, all work, including assembling the stiffness matrix and force vector and solving the equation system, is performed with our Fortran codes. It is noted that the stiffness matrix $\mathbf{K}_{\mathrm{A}}$ lacks the positive definiteness beyond the buckling, and thus the iterative solution does not converge to the minimum point of the potential energy. A simple way to remedy this problem is to replace $\mathbf{K}_{n+1}$ with $\mathbf{K}_{n+1} + \delta\mathbf{I}$, where $\mathbf{I}$ is the identity matrix, and δ is a positive number. The repeated replacement can ensure that the solution converges to the real one. The stiffness matrix $\mathbf{K}_{n+1}$ generally becomes positive definite after a few cycles of replacement, and the standard Newton-Raphson method can then be resumed. The choice of δ should ensure that $\mathbf{K}_{n+1} + \delta\mathbf{I}$ is positive definite. The ideal value of δ is a positive number that is slightly larger than the magnitude of the most negative eigenvalue of $\mathbf{K}_{n+1}$ [27] because the larger the value of δ, the slower the convergence of the solutions. Therefore, to achieve a good convergence rate, we can first calculate the eigenvalue of $\mathbf{K}_{n+1}$ at each iterative step. However, the calculation of this eigenvalue consumes additional computational time. Therefore, for small-sized structures, the value of δ can be chosen by calculating the eigenvalue of $\mathbf{K}_{n+1}$. But, for large-scale structures, it may be determined by frequent attempts.

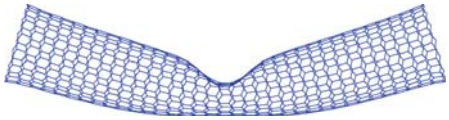
Using this atomic method, we have modeled the buckling behavior of SWCNTs under axial compression, twisting and bending (see Figure 14-4). The method is



(a) Axial compressive buckling of a (18,0) SWCNT with the length 8.72 nm.



(b) Twisting buckling of a (10,10) SWCNT with the length 11.81 nm.



(c) Bending buckling of a (15,0) SWCNT with the length 7.65 nm.

Figure 14-4. The buckling deformations of SWCNT under the axial compression, twisting and bending obtained by the present atomic simulation

proved to be much faster than the molecular dynamic. Prior to the buckling, several iterative steps can achieve a good convergence. Around the buckling, 100–200 iterative steps can generally give the stable results.

14.5. A HIGHER-ORDER CONTINUUM MODEL

An accurate multiscale scheme also requires a rational and efficient continuum model. However, it seems that the previous multiscale analyses always neglect the importance of the continuum models. Recently, the present authors employed the higher-order continuum theory to study CNTs, and good efficiency were obtained. The work is mainly motivated by the previous applications of the Cauchy-Born rule [14, 28] in the studies of CNTs. The Cauchy-Born rule is a fundamental kinematic assumption to bridge the deformation of the lattice vectors of crystal to that of a continuum deformation field, and it plays an important role in establishing the constitutive model. The Cauchy-Born rule has been used in the study of CNTs [29, 30]. However, this application may be insufficient and inaccurate. First, the Cauchy-Born rule requires sufficiently homogeneous deformations, but the deformation from a sheet to a 3-D curved surface is quite special, and the Cauchy-Born rule only maps a planar vector onto the tangent plane of the curved surface. Moreover, the computations of Arroyo and Belytschko [31–34] revealed that the constitutive model based

on the Cauchy-Born rule does not describe the bending effect, and thus the buckling deformation of CNTs cannot truly be displayed. The classical Cauchy-Born rule can be enhanced by considering the effect of the second-order deformation gradient, and this extended rule is called the higher-order Cauchy-Born rule [35–38]. With the higher-order Cauchy-Born rule, a higher-order constitutive model can be developed for SWCNTs [37, 38], and thus CNTs can be studied in the theoretical scheme of the higher-order gradient continuum [39, 40]. This section introduces the higher-order continuum model of SWCNTs and its mesh-free implementation.

14.5.1. Higher-Order Gradient Continuum

The essential idea of the higher-order gradient continuum is that the strain energy density depends not only on the first-order deformation gradient, but also on the second-order deformation gradient. As shown in Figure 14-5, we consider a curved surface in a 3-D space with its original image being a planar sheet on a 2-D plane. Here, we use the capital $\mathbf{X} = (X_1, X_2)$ to denote the original reference configuration and the lowercase $\mathbf{x} = (x_1, x_2, x_3)$ to denote the current configuration. The deformation map from the planar sheet to the curved surface is defined by $\mathbf{x} = \mathbf{x}(\mathbf{X})$, and the first- and second-order deformation gradients are

$$\mathbf{F}(i,J) = \frac{\partial \mathbf{x}}{\partial \mathbf{X}}, \ \mathbf{G}(i,J,K) = \frac{\partial^2 \mathbf{x}}{\partial \mathbf{X}^2}, \tag{14-22}$$

where i=1,2,3, J,K=1,2.

The first-order Piola-Kirchhoff stress tensor $\mathbf{P}$ and higher-order stress tensor $\mathbf{Q}$ are given by

$$\mathbf{P} = \frac{\partial W_0}{\partial \mathbf{F}}, \ \mathbf{Q} = \frac{\partial W_0}{\partial \mathbf{G}}, \tag{14-23}$$

where W_0 is the strain energy density that is defined in the reference configuration. The tangential moduli are given by

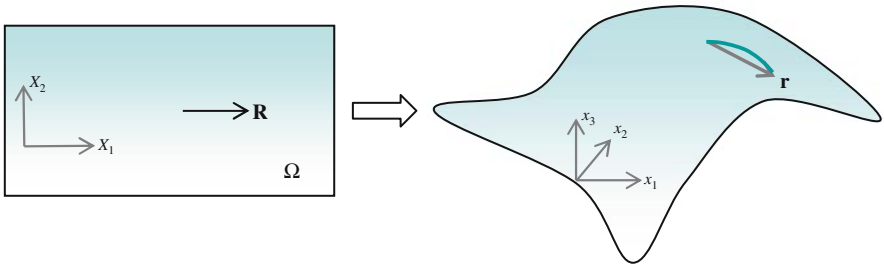


Figure 14-5. A planar sheet is transformed to a spatial curved surface. With this transformation, the vector in the original reference configuration is deformed as a curved line that corresponds to a spatial vector

$$\mathbf{M_{FF}} = \frac{\partial^2 W_0}{\partial \mathbf{F} \otimes \partial \mathbf{F}}; \mathbf{M_{FG}} = \frac{\partial^2 W_0}{\partial \mathbf{F} \otimes \partial \mathbf{G}}; \mathbf{M_{GF}} = \frac{\partial^2 W_0}{\partial \mathbf{G} \otimes \partial \mathbf{F}}; \mathbf{M_{GG}} = \frac{\partial^2 W_0}{\partial \mathbf{G} \otimes \partial \mathbf{G}}. \tag{14-24}$$

The stable configurations of the system are identified with the minimization of the total potential energy [36]:

$$E = \int_\Omega W_0(\mathbf{F},\mathbf{G})dV - \int_{\partial\Omega} \mathbf{u} \cdot \mathbf{t}_0^P dS - \int_{\partial\Omega} \nabla_N \mathbf{u} \cdot \mathbf{t}_0^Q dS, \tag{14-25}$$

where $\mathbf{t}_0^P$ and $\mathbf{t}_0^Q$ are, respectively, the first- and second-order stress tractions on the surface $\partial\Omega$ of the domain Ω the trial deformation $\mathbf{u} = (u_1,u_2,u_3)^{\mathrm{T}}$ must satisfy the essential boundary condition on the boundary of the domain. The outward normal gradient is defined as

$$\nabla_N \mathbf{u} = N_1 \frac{\partial \mathbf{u}}{\partial X_1} + N_2 \frac{\partial \mathbf{u}}{\partial X_2}, \tag{14-26}$$

with N_1 and N_2 being the unit outward normal components that are measured in the reference configuration.

With the deformation from a 2-D sheet to a curved surface in a 3-D space, an arbitrary vector $\mathbf{R}$ in the reference configuration is transformed to a curved line in the current configuration (see Figure 14-5), and this curved line corresponds to

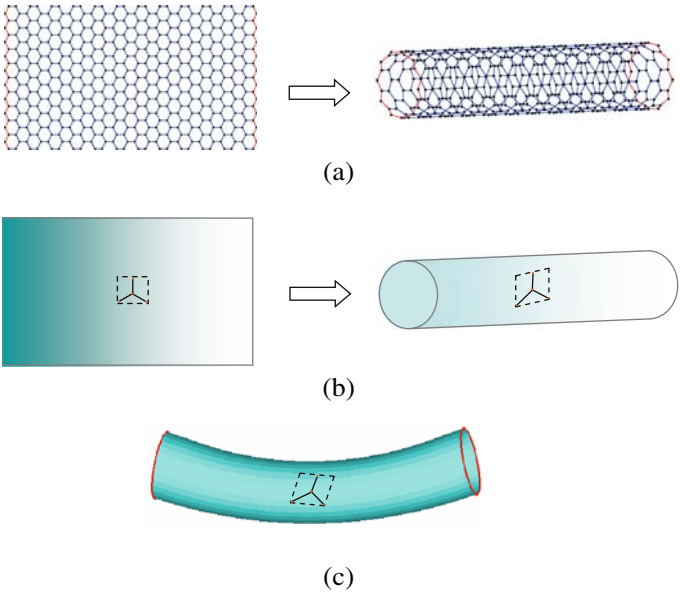


Figure 14-6. A graphite sheet is viewed as a continuum plane, and an SWCNT is viewed as a hollow cylindrical tube with a thin smooth surface. An undeformed SWCNT is formed by rolling up a planar sheet into a cylindrical shape. (a) Atomic structure; (b) continuum view; (c) the deformed SWCNTs

a special vector **r** that looks like a chord on the curved surface. For an arbitrary deformation, the exact evaluation of the vector **r** is difficult or impossible.

14.5.2. Constitutive Relationship

An undeformed SWCNT can be viewed as having been formed by rolling up a graphite sheet into a cylindrical shape (see Figure 14-6a). In the continuum model, a graphite sheet is viewed as a continuum plane, and an undeformed SWCNT is viewed as a hollow cylindrical tube with a thin smooth surface (Figure 14-6b). To evaluate the strain energy density at a given point, a representative cell structure is imagined at this point, as shown in Figure 14-6b,c (the orientation of the cell structure in the reference configuration is determined by the chirality of SWCNTs). A magnified cell structure is shown in Figure 14-7 in which the atom I is connected to bonds $I - J$ ($J = 1,2,3$). The deformation of the bond vectors is approximated using the higher-order Cauchy-Born rule [35–38].

$$\mathbf{r}_{IJ} = \mathbf{F} \bullet (\mathbf{R}_{IJ} + \boldsymbol{\eta}) + \mathbf{G}\text{:}[(\mathbf{R}_{IJ} + \boldsymbol{\eta}) \otimes (\mathbf{R}_{IJ} + \boldsymbol{\eta})]/\mathbf{2}, \tag{14-27}$$

here, $\boldsymbol{\eta} = (\eta_1,\eta_2)$ is the inner shift due to non-centrosymmetry of the atomic structure [14, 29–32].

The strain energy at this point can be calculated as

$$\hat{W}_0 = \hat{W}_0(\mathbf{F},\mathbf{G},\boldsymbol{\eta}) = \frac{V_I}{\Omega_I} = \frac{1}{2\Omega_I}\sum_{J=1}^{3} V_B(\mathbf{r}_{IJ}), \tag{14-28}$$

where V_B is the Brenner potential, Ω_I is the average area per atom in the reference configuration, and is calculated as $3\sqrt{3}r_0^2/4$ with r_0 being the bond length.

For a given macroscopic deformation description (**F** and **G**), the strain energy should be minimized with respect to the inner shift, which corresponds to

$$\frac{\partial \hat{W}_0}{\partial \boldsymbol{\eta}} = 0. \tag{14-29}$$

After the minimization, we obtain the equilibrium inner shift $\boldsymbol{\eta} = \boldsymbol{\eta}(\mathbf{F},\mathbf{G})$, which implies that the strain energy is only a function of the deformation gradients **F** and **G**:

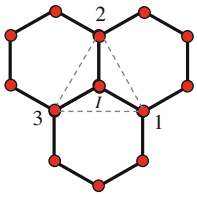


Figure 14-7. A representative cell composed of the bonds $I - J$ ($J = 1,2,3$)

$$W_0(\mathbf{F},\mathbf{G}) = \hat{W}_0(\mathbf{F},\mathbf{G},\eta) = \hat{W}_0(\mathbf{F},\mathbf{G},\eta(\mathbf{F},\mathbf{G})). \tag{14-30}$$

The Piola-Kirchhoff stress tensor **P** and **Q** can be calculated using Eq. (14-23), and the tangential moduli can be calculated using Eq. (14-24).

14.5.3. Mesh-Free Numerical Simulation

Based on the above higher-order constitutive model, we have developed a mesh-free computational scheme for SWCNTs in Refs. [39–42]. Employing the planar sheet as the reference configuration (see Figure 14-6b), the deformation from the reference configuration to the current configuration can be decomposed into two parts. The first part is from the reference configuration to the initial equilibrium SWCNT, and it can be exactly calculated. Another part is from the initial equilibrium SWCNT to the current configuration, and this part is treated as unknown. In the mesh-free simulation, the second part is interpolated with the moving least-square approximation [43–46]. For any evaluated point, the first- and second-order deformation gradients are interpolated directly with the nodal parameters. Of course, we need to use the first- and second-order derivatives of the mesh-free shape function. The stable configurations can thus be solved using the Newton's method [21].

Using the present mesh-free method, we have carried out the numerical simulations of SWCNTs under the axial compression and twisting. At the same time, we also carried out numerical simulation based on the classical Cauchy-Born rule [23, 24]. Figure 14-8 shows a comparison of the buckling patterns obtained with two

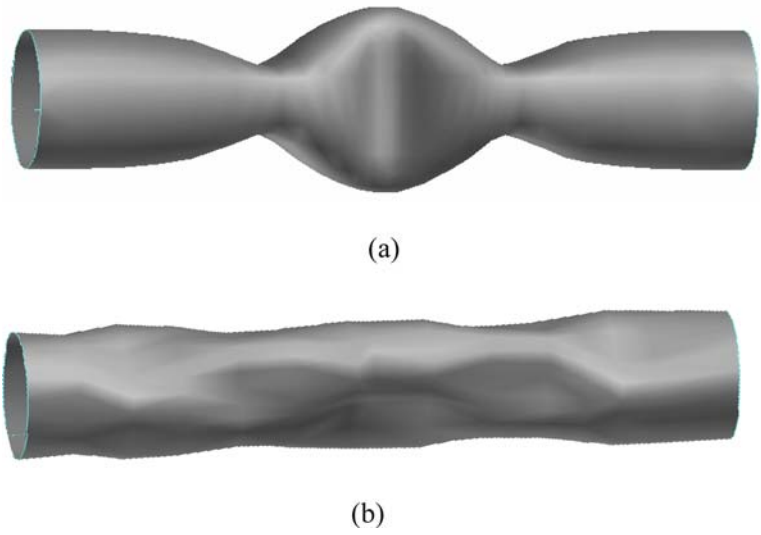


Figure 14-8. A comparison of the axial compressive buckling deformations of a (18,0) SWCNT with the length 8.72 nm obtained with (**a**) the higher-order continuum model; and (**b**) the classical Cauchy-Born rule

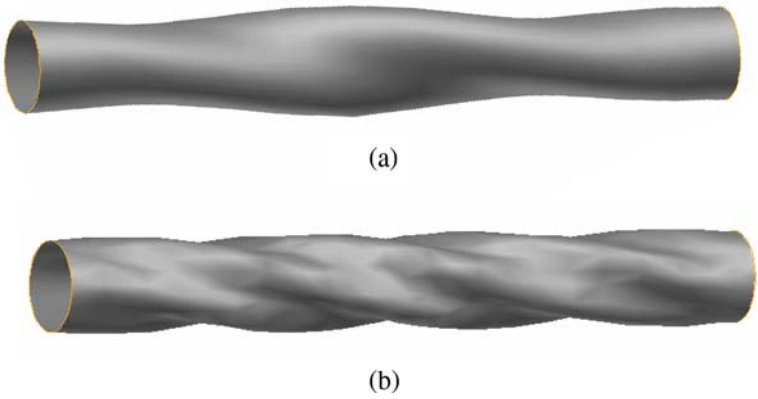


Figure 14-9. A comparison of the twisting buckling deformations of a (10,10) SWCNT with the length 11.81 nm obtained with (**a**) the higher-order continuum model; and (**b**) the classical Cauchy-Born rule

continuum models. Obviously, the higher-order continuum model can truly display the buckling deformation of SWCNTs, but the continuum model based on the classical Cauchy-Born rule results in a non-physical pattern. Shown in Figure 14-9 are the twisting buckling patterns obtained with the two continuum models. Figures 14-8b and 14-9b are the response of structures in case of no bending stiffness. In addition, our computations showed that, prior to the buckling, the mesh-free computational precision is very good. But, around the buckling, the mesh-free simulation becomes less accurate whereas an amount of nodes were used [39–42].

14.6. MULTISCALE COUPLING SCHEME

The present multiscale scheme couples the developed mesh-free method and the atomic simulation method presented in Section 14.5. The coupling approach used is the bridging domain method described in Section 14.2.

Referring to Figure 14-2, the entire domain is decomposed into three regions: Ω^{A}, $\Omega^{\mathrm{C}} = \Omega^{\mathrm{CL}} + \Omega^{\mathrm{CR}}$, and $\Omega^{\mathrm{O}} = \Omega^{\mathrm{A}} \cap \Omega^{\mathrm{C}}$. The mesh-free method is used in the continuum region. Mesh-free nodes are uniformly collocated along the axial and circumferential directions, and the background integral cells are in accordance with the nodal arrangements. The crossing points of the axial and circumferential lines denote the mesh-free nodes in Figure14-2.

The total energy of the coupled system can be written as a weighted sum of energies for the continuum and atomic regions:

$$\mathrm{U_{tot}} = \sum_{I \in \Omega^{\mathrm{A}}} \left[\beta^{\mathrm{A}} \left(\frac{1}{2} (\mathbf{q}_I^{\mathrm{A}} + \mathbf{q}_J^{\mathrm{A}}) \right) \sum_{I<J} V_B(\mathbf{q}_J - \mathbf{q}_I) \right] + \int_{\Omega^c} \beta^{\mathrm{C}}(\mathbf{X}) W_0(\mathbf{F},\mathbf{G}) dV, \tag{14-31}$$

where $\mathbf{q}_I^A$ is the position of atom I in the undeformed SWCNT, and the weight functions β^{A} and β^{C} take the forms [16, 17, 21]

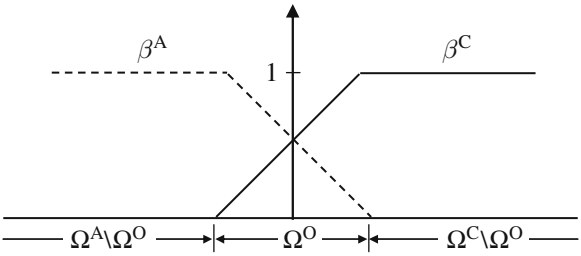


Figure 14-10. The definition of weight functions β^{A} and β^{C} in the three regions

$$\beta^{\mathrm{C}}(\mathbf{X}) = 1 - \beta^{\mathrm{A}}(\mathbf{X}) = \begin{cases} 1, & \mathrm{x} \in \Omega^{\mathrm{C}}\backslash\Omega^{\mathrm{A}}, \\ \zeta, & \mathrm{x} \in \Omega^{\mathrm{A}}, \\ 0, & \mathrm{x} \in \Omega^{\mathrm{A}}\backslash\Omega^{\mathrm{C}}, \end{cases} \tag{14-32}$$

where the symbol "\" denotes the set-minus operation, and the parameter ζ varies linearly from 0 to 1 across the overlapping region, as shown in Figure 14-10. This method thus allows the minimization of the continuum and atomic configurations concurrently, and the two scale solutions can simultaneously be solved.

The total potential energy of the coupled system is

$$E_{\mathrm{tot}} = U_{\mathrm{tot}} - \sum_{i \in \Omega^{\mathrm{A}}} \beta^{\mathrm{A}}(\mathbf{q}_I^{\mathrm{A}}) \bar{f}_I \cdot \mathbf{q}_I - \int_{\partial\Omega^{\mathrm{c}}} \beta^{\mathrm{C}}(\mathbf{X}) \mathbf{u} \cdot \mathbf{t}_0^P dS - \int_{\partial\Omega^{\mathrm{c}}} \beta^{\mathrm{c}}(\mathbf{X}) \nabla_N \mathbf{u} \cdot \mathbf{t}_0^Q dS. \tag{14-33}$$

The non-equilibrium force vector and stiffness matrix can be obtained by

$$\mathbf{f}_{\mathrm{tot}} = \frac{\partial E_{\mathrm{tot}}}{\partial \mathbf{X}_{\mathrm{tot}}}, \tag{14-34}$$

$$\mathbf{K}_{\mathrm{tot}} = \frac{\partial^2 E_{\mathrm{tot}}}{\partial \mathbf{X}_{\mathrm{tot}} \partial \mathbf{X}_{\mathrm{tot}}}. \tag{14-35}$$

In Refs. [16, 17], the augmented Lagrange multiplier method was used to ensure that the continuum displacements conformed to the atomic displacements at the discrete positions of the atoms in the overlapping region. In the present research, the atomic displacement in the overlapping region is directly interpolated with the mesh-free nodal parameters at each iterative step, which means that the vector $\mathbf{X}_{\mathrm{tot}}$ contains all of the mesh-free modal parameters, but only the current positions of the atoms outside the overlapping region. The stable configuration can be obtained by iteratively solving $\mathbf{K}_{\mathrm{tot}} \Delta \mathbf{X}_{\mathrm{tot}} = \mathbf{f}_{\mathrm{tot}}$ until $\mathbf{f}_{\mathrm{tot}}$ reaches zero. Moreover, the modification method of the stiffness matrix should also be used around the buckling.

14.7. MULTISCALE COMPUTATIONAL EXAMPLES

Two examples are chosen to test the validity of the present multiscale computational scheme. The first example is the bending test, for which the Brenner potential with the second set of parameters is used. The second example considers the fracture of SWCNTs with a single-atom vacancy. As described in Section 14.3, the Brenner potential always results in over-estimated fracture stress. Therefore, the modified second-generation Brenner potential is used for the second example.

14.7.1. Bending Test

A (12,12) SWCNT is chosen upon which to perform the bending test. The SWCNT has 42 hexagonal cells (a total of 2,040 atoms) in its length. Atomic simulation is used for the middle portion, whereas the left and right parts are modeled using a mesh-free method in which 396 mesh-free nodes are used. In the overlapping region, the atomic position is interpolated using the mesh-free nodes, and the pure atomic region contains 552 atoms. The degrees of freedom of the system are reduced from 3×2024 to 3×948. The bending deformation is imposed by incrementally rotating the two ends of SWCNT in the opposite direction. For each loading step, the tube is bent by 1° per loading step. Figure 14-11 shows a plot of the changes in the average

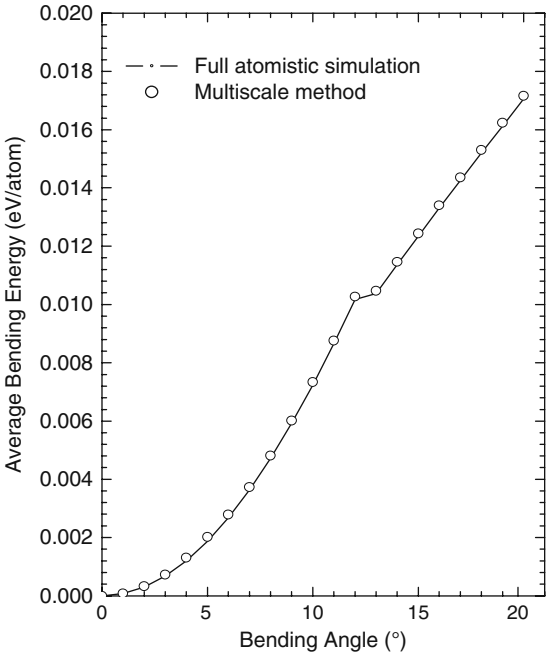


Figure 14-11. Comparison of the average bending energy derived from the multiscale method and the full atomic simulation

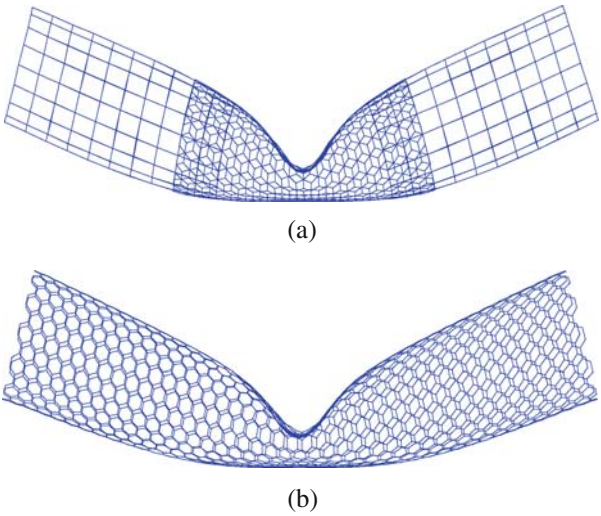


Figure 14-12. Deformation of the SWCNT at a bending angle of 20°, as obtained by (**a**) the multiscale method and (**b**) the full atomic simulation

energy per atom against the bending angle for the multiscale simulation and the full atomic simulation. Buckling occurs at a bending angle of 13° in both simulations, and the energy change from the two methods also agrees quite well both before and after the buckling occurs. Figure 14-12 shows the deformations at a bending angle of 20°, and a comparison of the two methods reveals that a precise simulation can be achieved with the multiscale method. In Section 14.6, we have indicated that the mesh-free simulation presents very precise results prior to buckling, but after buckling, the results become less accurate. The present multiscale simulation presents a good agreement with the atomic simulation before and after buckling, which shows that the multiscale method is very efficient for the localized problems.

14.7.2. Tensile Failure of SWCNTs with a Single-Atom Vacancy Defect

In this example, the multiscale method is used to study the tensile failure of an SWCNT with a single-atom vacancy defect. This kind of defect has a significant effect on the strength of a CNT and thus has attracted considerable research attention. In the computation, one atom is removed from the hexagonal network in the atomic region, which leads to the reconstruction of the atomic structure near the vacancy in the initial equilibrium configuration [15]. Here, this initial equilibrium configuration is determined with the full atomic simulation before the multiscale computation is performed. The SWCNT is stretched until the fracture occurs. Figure 14-13 shows the stress-strain curves for (18, 0) and (12, 12) SWCNTs. The tensile force can be computed by summing the nodal force along the end of the CNT, and the stress is calculated by dividing the tensile force by the cross-sectional area $2\pi Rt$ (t being the thickness of SWCNT, and chosen as 0.34 nm). Figure 14-14 shows the cracking procedure for an (18,0) SWCNT, with Figure 14-14a showing the initial

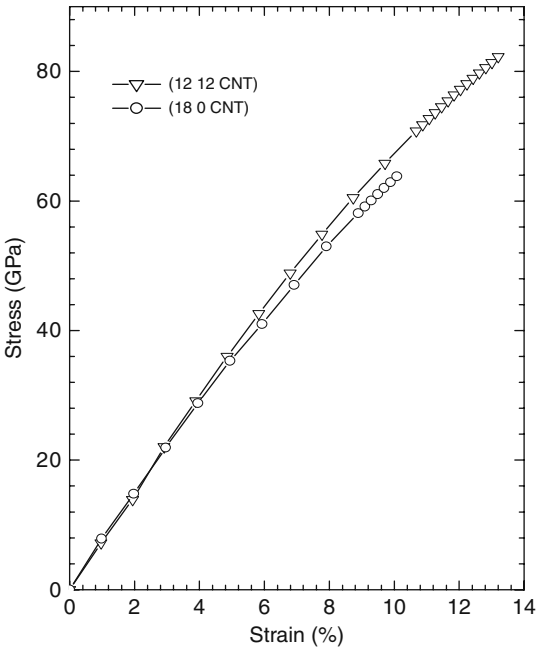


Figure 14-13. Stress-strain curves for the tension of (18,0) and (12,12) SWCNTs with a single-atom vacancy defect

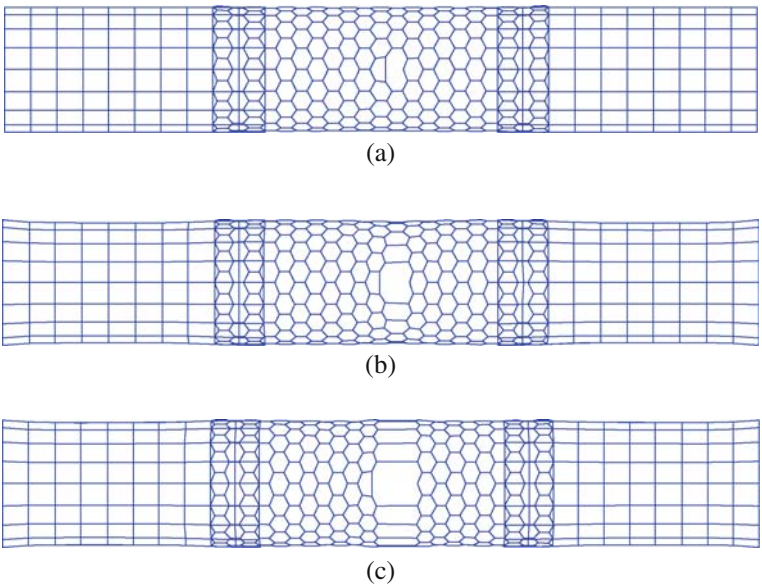


Figure 14-14. Fracture evolution of an (18,0) SWCNT with a single-atom vacancy defect: (**a**) The initial equilibrium configuration of the undeformed SWCNT; (**b**) bond failure spreading from the defect; and (**c**) the fractured SWCNT, in which all of the bonds around the circumference have failed

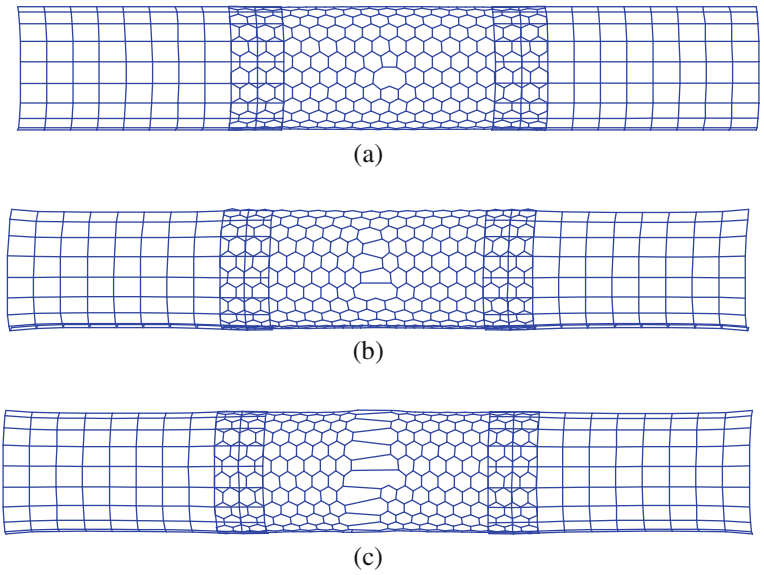


Figure 14-15. Fracture evolution of a (12,12) SWCNT with a single-atom vacancy defect: (**a**) the initial equilibrium configuration of the undeformed SWCNT; (**b**) bond failure spreading from the defect; and (**c**) the fractured CNT in which all of the bonds around the circumference have failed

equilibrium configuration of the undeformed SWCNT, Figure 14-14b highlighting the breakage of the bonds around the defect, and Figure 14-14c displaying the completely fractured SWCNT in which all of the bonds around the circumference have failed. Figure 14-15 shows the fracture progression for a (12,12) SWCNT, from which it can be seen that the failure path displays a slight difference from that of the (18,0) SWCNT. In the present simulation, the failure is actually a brittle fracture. To show the cracking progression, a larger positive number β is added to the diagonal elements of the stiffness matrix after the occurrence of the fracture.

14.8. SUMMARY

Due to the limitation of the atomic simulation and continuum methods, the multiscale modeling is emerging as a feasible and efficient approach for large-size nanostructures. This work presents a comprehensive understanding for the multiscale modeling of CNTs by illustrating all three components: atomic simulation, continuum modeling, and multiscale coupling scheme. An atomic simulating method that is faster than MD is introduced. Multiscale modeling of CNTs requires a reasonable and rational continuum model. A higher-order continuum model is introduced, and mesh-free method is employed to implement the continuum discretization. The coupling between the mesh-free method and the atomic simulation is achieved with the bridging domain method. The implementation of multiscale modeling of CNTs is discussed in detail.

REFERENCES

1. Iijima S (1991) Helical microtubules of graphitic carbon. Nature (London) 354: 56–58
2. Yakobson BI, Brabec CJ, Bernholc J (1997) Nanomechanics of carbon tubes: Instabilities beyond linear response. Phys Rev Lett 76: 2511–2514
3. Liew KM, He XQ, Wong CH (2004) On the study of elastic and plastic properties of multi-walled carbon nanotubes under axial tension using molecular dynamics simulation. Acta Materialia 52: 2521–2527
4. Liew KM, Wong CH, Tan MJ (2005) Buckling properties of carbon nanotube bundles. Appl Phys Lett 87: 041901
5. Govindjee S, Sackman JL (1999) On the use of continuum mechanics to estimate the properties of nanotubes. Solid State Commun 110: 227–230
6. Ru CQ (2000) Effective bending stiffness of carbon nanotubes. Phys Rev B 62: 9973–9976
7. Ru CQ (2001) Axially compressed buckling of a double-walled carbon nanotubes embedded in an elastic medium. J Mech Phys Solids 49: 1265–1279
8. Li CY, Chou TW (2003) A structural mechanics approach for the analysis of carbon nanotubes. Int J Solids Struct 40: 2487–2499
9. Li CY, Chou TW (2004) Elastic properties of single-walled carbon nanotubes in transverse directions. Phys Rev B 69: 073401
10. He XQ, Kitipornchai S, Liew KM (2005) Buckling analysis of multi-walled carbon nanotubes: A continuum model accounting for van der Waals interaction. J Mech Phys Solids 53: 303–326
11. Liew KM, He XQ, Kitipornchai S (2005) Buckling characteristics of embedded multi-walled carbon nanotubes. Proc R Soc Lond Ser A, Math Phys Engng Sci 46: 3785–3805
12. Wang LF, Hu HY (2005) Flexural wave propagation in single-walled carbon nanotubes. Phys Rev B 7: 195–412
13. Wang LF, Hu HY, Guo WL (2006) Validation of the non-local elastic shell model for studying longitudinal waves in single-walled carbon nanotubes. Nanotechnology 17: 1408–1415
14. Tadmor EB, Smith GS, Bernstein N, Kaxiras E (1999) Mixed finite element and atomistic formulation for complex crystals. Phys Rev B 59: 235–245
15. Tadmor EB, Ortiz M, Phillips R (1996) Quasicontinuum analysis of defects in solids. Philos Mag A 73(6): 1529–1563
16. Belytschko T, Xiao SP (2003) Coupling methods for continuum model with molecular model. Int J Multiscale Comput Engng 1: 115–126
17. Xiao SP, Belytschko T (2004) A bridging domain method for coupling continua with molecular dynamics. Comput Methods Appl Mech Engng 193: 1645–1669
18. Qian D, Wagner GJ, Liu WK (2004) A multiscale projection method for the analysis of carbon nanotubes. Comput Methods Appl Mech Engng 193: 1603–1632
19. Wagner GJ, Liu WK (2003) Coupling of atomistic and continuum simulations using a bridging scale decomposition. J Comput Phys190: 249–274
20. Shenoy VB, Miller RE, Tadmor EB, Rodney D, Phillips R, Ortiz M (1999) An adaptive finite element approach to atomic-scale mechanics-quasicontinuum method. J Mech Phys Solids 47: 611–642
21. Zhang SL, Mielke SL, Khare R, Troya D, Ruoff RS, Schatz GC, Belytschko T (2005) Mechanics of defects in carbon nanotubes: Atomistic and multiscale simulations. Phys Rev B 71:115403
22. Tersoff J (1988) New empirical approach for the structure and energy of covalent systems. Phys Rev B 37: 6991–7000
23. Brenner DW (1990) Empirical potential for hydrocarbons for use in simulating the chemical vapour deposition of diamond films. Phys Rev B 42(15): 9458–9471

24. Brenner DW, Shenderova OA, Harrison JA, Stuart SJ, Ni B, Sinnott SB (2002) A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. J Phys: Condensed Matter 14: 783–802
25. Liu B, Huang Y, Jiang H, Qu S, Hwang KC (2004) The atomic-scale finite element method. Compu Methods Appl Mech Engng 193: 1849–1864
26. Liu B, Jiang H, Huang Y, Qu S, Yu MF, Hwang KC (2005) Atomic-scale finite element method in multiscale computation with applications to carbon nanotubes. Phys Rev B 72: 035435
27. Dennis JE, Schnabel RB (1983) Numerical methods for unconstrained optimization and nonlinear equations. Prentice-Hall Inc, New Jersey
28. Weiner JH (1983) Statistical mechanics of elasticity. Wiley, New York
29. Zhang P, Huang Y, Geubelle PH, Klein PA, Hwang KC (2002) The elastic modulus of single-wall carbon nanotubes: A continuum analysis incorporating interatomic potentials. Int J Solids Struct 39: 3893–3906
30. Zhang P, Jiang H, Huang Y, Geubelle PH, Hwang KC (2004) An atomistic-based continuum theory for carbon nanotubes: Analysis of fracture nucleation. J Mech Phys Solids 52: 977–998
31. Arroyo M, Belytschko T (2002) An atomistic-based finite deformation membrane for single layer crystalline films. J Mech Phys Solids 50: 1941–1977
32. Arroyo M, Belytschko T (2004) Finite element methods for the non-linear mechanics of crystalline sheets and nanotubes. Int J Numer Methods Engng 59: 419–456
33. Arroyo M, Belytschko T (2005) Continuum mechanics modeling and simulation of carbon nanotubes. MECCANICA, 40: 455–469.
34. Arroyo M, Belytschko T (2004) Finite element methods for the non-linear mechanics of crystalline sheets and nanotubes. Int J Numer Methods Engng 59: 419–456
35. Leamy MJ, Chung PW, Namburu R (2003) On an exact mapping and a higher-order Born rule for use in analyzing graphene carbon nanotubes. In: Proceedings of the 11th Annual ARL-USMA Technical Symposium, Aberdeen
36. Sunyk R, Steinmann P (2003) On higher gradients in continuum-atomistic modeling. Int J Solids Struct 40: 6877–6896
37. Guo X, Wang JB, Zhang HW (2006) Mechanical properties of single-walled carbon nanotubes based on higher order Cauchy-Born rule. Int J Solids Struct 43: 1276–1290
38. Wang JB, Guo X, Zhang HW, Wang L, Liao JB (2006) Energy and mechanical properties of single-walled carbon nanotubes predicted using the higher order Cauchy-Born rule. Phys Rev B 73: 115428
39. Sun YZ, Liew KM (2008) Mesh-free simulation of single-walled carbon nanotubes using higher order Cauchy-Born rule. Comput Mater Sci 42(3): 444–452
40. Sun YZ, Liew KM (2008) The buckling of single-walled carbon nanotubes upon bending: The higher order gradient continuum and mesh-free method. Comput Methods Appl Mech Engng197: 3001–3013
41. Liew KM, Sun YZ (2008) Elastic properties and pressure-induced structural transitions of single-walled carbon nanotubes. Phys Rev B 77: 205437
42. Sun YZ, Liew KM (2008) Application of the higher-order Cauchy-Born rule in mesh-free continuum and multiscale simulation of carbon nanotubes. Int J Numer Methods Engng, 75(10): 1238–1258.
43. Belytschko T, Krongauz Y, Organ D, Fleming M, Krysl P (1996) Meshless methods: An overview and recent developments. Compu Methods Appl Mech Engng 139: 3–47
44. Liew KM, Lim HK, Tan MJ, He XQ (2002) Analysis of laminated composite beams and plates with piezoelectric patches using the element-free Galerkin method. Comput Mech 29: 486–497
45. Liew KM, Huang YQ, Reddy JN (2003) Moving least square differential quadrature method and its application to the analysis of shear deformable plates. Int J Numer Methods Engng 56: 2331–2351
46. Tang Z, Shen S, Atluri SN (2004) Analysis of materials with strain-gradient effects: A meshless local Petrov-Galerkin (MLPG) with nodal displacements only. Comput Modeling Engng Sci 4: 177–196

CHAPTER 15

QUASICONTINUUM SIMULATIONS OF DEFORMATIONS OF CARBON NANOTUBES

SEYOUNG IM[1], SUNGJIN KWON[1], AND JONG YOUN PARK[2]

[1]*Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Science Town, Daejeon, Korea, e-mail: sim@kaist.ac.kr*

[2]*Technical Research Laboratory POSCO, Pohang, Gyeongbuk, Korea*

Correspondance to S. Im, Sim@kaist.ac.kr

Abstract: The noble computational scheme known as quasicontinuum (QC) has been widely utilized over the past decade for exploring extreme/multi-scale phenomena in the spatial domain, such as, mechanical behaviors of nanostructures or defect behaviors in crystalline materials. This article reports on the recent extension of the QC method to simulate mechanical behaviors or deformations of curved crystalline bodies such as carbon nanotubes (CNTs). In addition to QC implementation utilizing high-order triangular elements, this study presents a new QC approach based on what is known as "variable-node elements". This proves to be extremely efficient when combined with a fully automatic adaptive refinement. Several numerical examples demonstrate the accuracy and effectiveness of the new method.

Keywords: Quasicontinuum, Multiscale computation, Carbon nanotubes

15.1. INTRODUCTION

Coarse-graining schemes have been widely utilized as a useful tool for bridging the gap between continuum modeling and discrete atomistic modeling. They make it possible to look into atomistic material behaviors in consideration of length scales greater than atomic or molecular scales. Among others, the quasicontinuum (QC) method, first reported by Tadmor et al. [1], has been highly successful in exploring the behaviors of defects such as dislocations, grain boundaries, twins, and voids and impurities (see [2] for details).

Tadmor et al. [1] effectively differentiated between the local zone and the nonlocal zone, providing a corresponding formulation appropriate for each of both zones and coupling them with each other to consider the entire zone. The local zone here implies the part of the domain of a body where the Cauchy-Born rule is applicable, whereas the nonlocal zone refers to the part of the body for which the Cauchy-Born

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 389–419.
DOI 10.1007/978-1-4020-9785-0_15,

rule is not applicable. In the context of continuum mechanics, locally homogeneous (or uniform) elastic deformations prevail in the local zone, so that the energy of a local subdomain, which is mostly an element, is obtained by multiplying the energy of one atom by the number of atoms contained in the element or by multiplying the strain energy density by the element volume. The calculation of the energy per unit volume or per atom may be conducted in various ways depending on the nature of the simulation (see [2] for detail). On the other hand, the presumption of the locally homogeneous elastic deformations is not valid in the nonlocal zone; here, it is necessary to calculate the energy directly over a nonlocal subdomain, which is in practice mostly an individual atom or an element, from the discrete atomistic viewpoint.

Each of the two formulations, the local and the nonlocal approach, has its own advantages. The local approach provides an extremely efficient means of computation through a coarse-graining scheme based on Cauchy-Born rule or the locally homogeneous elastic deformations. This assumption makes it possible to calculate the strain energy density function for the local subdomain either from the atomistic viewpoint or from the continuum viewpoint by calculating the energy of a unit cell or one atom. Subsequently, an explicit expression for the stress can be obtained by differentiating the strain energy density function. This does not mean that the local formulation is merely the treatment according to the theory of finite elasticity, which utilizes a phenomenological strain energy density function. If the local approach calculates the strain energy density function from the atomistic viewpoint, as is often done, it preserves the crystal orientation and the corresponding anisotropy. On the other hand, the nonlocal approach sees the surroundings beyond the local neighborhood. Therefore, simple homogenization is not allowed, but the nonlocal subdomain energy should be coarse-grained through the use of the discrete atomic energy for an individual atom. As a consequence, the nonlocal formulation is capable of providing atomic scale resolution. Thus, it properly describes the abrupt changes or disturbances of atomic behaviors in crystalline bodies due to defects such as dislocations, grain boundaries, twins or free surfaces.

An appropriate combination of the local and the nonlocal formulation leads to a coupled approach. In this approach, the nonlocal QC is applied in regions where the gradient of strain is large, or on nonlocal regions to capture the discrete atomistic behaviors. In contrast, the local QC is applied in regions where deformation is relatively uniform or in local regions for an effective reduction of the degrees of freedom. Hence, the coupled QC appears to provide a computational scheme that bridges the local zone and the nonlocal zone seamlessly. However, some inconsistency may be found between the local formulation and the nonlocal formulation, so that the "ghost force" may take place [3]. Knap and Ortiz [4] proposed the purely nonlocal approach known as the "cluster-based summation rule," which sums the energy contribution of every atom within a given cluster associated with each representative node, instead of the "node-based summation rule." This fully nonlocal version paves the way to a seamless coupling of different length scales without any possibility of the inconsistency such as the ghost force. In addition, this method eliminates any possible zero energy modes that may be found in the node-based summation rule [4].

Among the key ingredients of the QC is an effective adaptive meshing scheme that controls the sizes of the individual element subdomains in an adaptive manner, consistent with the severity of the deformation and the deformation gradient. This scheme enables one to reach optimal meshing or coarsening so that a high resolution, very often down to the atomic scale, may be obtained in regions of a stiff gradient of deformation, while appropriate reduction of the degrees of freedom may be accomplished in regions of relatively uniform deformation. The element refinement or coarsening depends on the deformation and deformation gradient, which is scarcely known a priori in many nanomechanics problems. In QC applications for such complex problems, accordingly, an adaptive meshing or refinement scheme is essential for an efficient and accurate solution. Mostly, adaptive refinement is implemented [3] by use of linear triangular finite elements which are generated utilizing the scheme of the constrained Delaunay triangulation [5]. In practice, most of the finite-element implementations for the QC in the literature are limited to linear triangular or tetrahedron elements, as adaptive refinement or meshing is achieved most readily through triangulation for this type of elements.

One noteworthy point regarding the use of the linear tetrahedron or triangular elements for the QC is that applications of the QC have been severely restricted to rectilinear crystalline structures or materials. Hence, applications for curved bodies like CNTs (carbon nanotubes) are rare apart from the recent work by Park and Im [6]. This is the natural consequence of using linear tetrahedron triangular elements, as these elements are not capable of modeling curved geometries. In addition, despite the high geometric adaptability of triangular or tetrahedron elements, these elements are far from being satisfactory in terms of the solution accuracy and convergence rate. In terms of the availability of an efficient alternative for adaptive refinement, it is not necessary to use only triangular or tetrahedral elements for the implementation of the QC.

The purpose of this chapter is to provide a wider view regarding the implementation of the QC so as firstly to extend its application range to include curved crystalline bodies such as carbon nanotubes (CNTs) and secondly to introduce a new adaptive quasicontinuum implementation for CNTs using what are known as variable-node elements. The outline of this chapter is as follows. In Section 15.2, simulation of deformation behaviors of CNTs, as typical curved crystalline structures, is reviewed from the view of a multiscale computation, and a nonlocal QC scheme along the line of Park and Im [6] is discussed with a view to an application of the simulation for CNT deformations. Next, in Section 15.3 a new adaptive QC scheme based on variable-node elements [7–9] is proposed to extend the QC method to rectangular elements other than tetrahedron or triangular elements. This is followed by Section 15.4, which concludes this chapter.

15.2. QUASICONTINUUM METHOD FOR CARBON NANOTUBES

There have been extensive studies of the simulation of CNT deformation from the viewpoint of continuum or structural mechanics as well as from the viewpoint of an atomistic approach (for example, see [6, 10–12] and the references cited therein).

Among others, Arroyo and Belytschko [11] utilized the exponential Cauchy-Born rule to extend the standard Cauchy-Born rule to the case of curved crystalline bodies. This enabled them to devise an efficient coarse-graining scheme for the local quasicontinuum approximation to deformations of curved crystalline structures such as CNTs.

The first multiscale approach to the nonlocal deformation behaviors of CNTs was reported by Qian et al. [12], where a bridging scheme between continuum and atomistic calculations is employed to look into CNT deformations down to the atomic scale. This is a useful approach for multiscale computations of CNT deformations; however, no adaptive meshing or refinement was employed in the computation, and its applicability is somewhat limited in this context. Typically, some complex multiscale problems are encountered for which it is not known a priori where zones of stiff deformation gradient or nonlocal zones are to occur. For this category of problems, it is essential to couple the computing scheme with an appropriate adaptive refinement or meshing strategy.

Recently, Park and Im [6] generalized the nonlocal QC of Knap and Ortiz [4], which was applied exclusively for rectilinear crystalline structures, for application to curved crystalline bodies such as CNTs. Their work combines the nonlocal QC scheme with a fully automatic adaptive refinement method, providing an effective multiscale computing tool for CNT deformations and enabling one to look into nonlocal regions with resolution down to the atomic scale as well as to coarsen local regions in an effective manner. The main feature of this methodology includes the use of high-order interpolation functions for the accurate mapping of curved CNT geometry. This was made possible through the introduction of the concept of atomless nodes [6] into the nonlocal formulation of Knap and Ortiz, which is based on the cluster-based summation rule [4].

In this section, an earlier study [6] is reviewed, extending the nonlocal QC method of Knap and Ortiz [4] to the case of curved crystalline bodies such as CNTs. This is a fully nonlocal treatment that makes it possible to look into the atomistic behaviors of CNTs as well as into the continuum-scale deformations of CNTs.

15.2.1. Deformations of Single-Walled CNTs

As in the mechanics of continuous media, a reference configuration is needed to describe the deformations of a CNT. The initial state is often the choice of a reference configuration, as in the total Lagrangian description for nonlinear finite element analysis for continua. In an atomistic computation of nano-systems or nanostructures, however, the initial configuration itself is not a state completely known a priori in the presence of boundary layers such as free surfaces or interfaces. The initial state of a CNT should be obtained through energy minimization as the exact initial atomic configuration, particularly including the free ends, may be substantially away from the bulk state of the graphene due to bond breakage or curvature effects. One convenient choice of a reference configuration for CNTs is the bulk state of the graphene, which is known a priori.

Consider a single-walled CNT (SWCNT). Here, it is assumed that the CNT undergoes a locally homogeneous deformation in which the Cauchy-Born rule (or

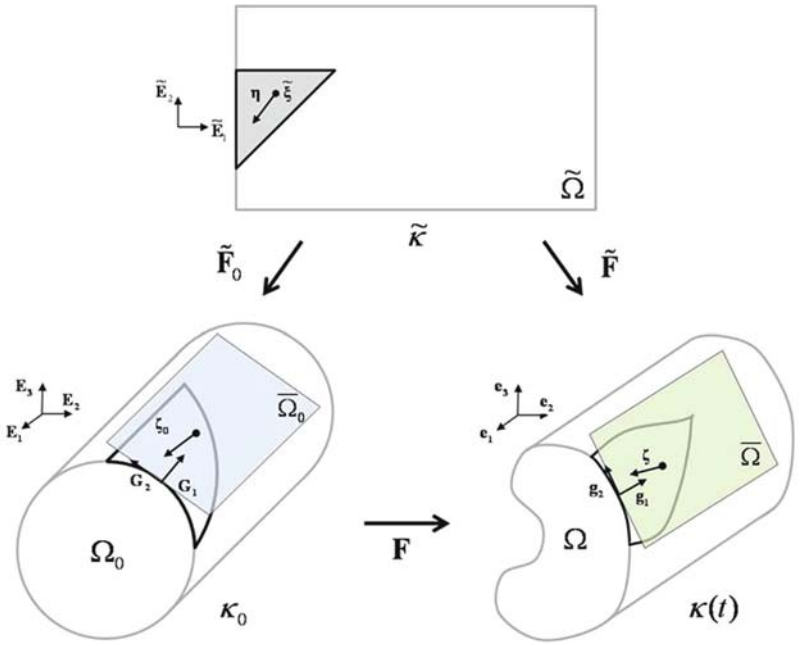


Figure 15-1. Deformations of a CNT and three configurations and$\tilde{\kappa}$, κ_o and $\kappa(t)$

exponential Cauchy-Born rule, strictly speaking) is valid so that lattice is deformed according to a locally uniform tensor of deformation gradient. The CNT geometry is specified by the chirality and the number of carbon layers along the axial direction. Energy minimization yields its radius and length, and the geometry of the initial configuration κ_o is fully determined (see Figure15-1). For the reference configuration, the bulk state of the flat graphene sheet $\tilde{\kappa}$ is chosen, as shown in Figure 15-1.

Two deformations are then considered: one from the bulk graphene state $\tilde{\kappa}$ to the initial configuration κ_o, and the other from $\tilde{\kappa}$ to the current deformed configuration $\kappa(t)$, as shown in Figure 15-1. Let $\tilde{\mathbf{Y}} \epsilon \mathbb{R}^2$,$\mathbf{X} \epsilon \mathbb{R}^3$, and $\mathbf{x} \epsilon \mathbb{R}^3$ denote the position vectors in $\tilde{\kappa}$, κ_o and $\kappa(t)$, respectively. The aforementioned deformation may be written as

$$\mathbf{X} = \tilde{\mathbf{\Phi}}_o(\tilde{\mathbf{Y}}) \text{ and } \mathbf{x} = \tilde{\mathbf{\Phi}}(\tilde{\mathbf{Y}}) \tag{15-1a,b}$$

$$\mathbf{x} = \tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}_o^{-1}(\mathbf{X}) = \mathbf{\Phi}(\mathbf{X}) \tag{15-1c}$$

where $\tilde{\mathbf{\Phi}}_o$,$\tilde{\mathbf{\Phi}}$ and $\mathbf{\Phi}$ indicate the mappings associated with the corresponding deformations. Let $\tilde{\mathbf{F}}_o$ and $\tilde{\mathbf{F}}$ indicate the deformation gradients corresponding to the two deformations in Eq. (15-1a,b):

$$\tilde{\mathbf{F}}_o = \nabla_{\tilde{\mathbf{Y}}}\tilde{\mathbf{\Phi}}_o \text{ and } \tilde{\mathbf{F}} = \nabla_{\tilde{\mathbf{Y}}}\tilde{\mathbf{\Phi}} \tag{15-2a,b}$$

Then the deformation gradient $\mathbf{F}$ from κ_o to $\kappa(t)$ is given as

$$\mathbf{F} = \nabla_{\mathbf{X}} \mathbf{\Phi} = \tilde{\mathbf{F}} \tilde{\mathbf{F}}_o^{-1} \tag{15-3}$$

Several assumptions on deformation kinematics as in elastic shells enable a description of CNT deformations in terms of the extension and the curvature on the midplane. For local formulation in consideration of atomic features such as anisotropy and the crystal orientation, these strain measures may directly be linked to the interatomic potentials of CNTs. This will then result in an effective local coarse-graining scheme, such as that by Arroyo and Belytschko [11]. This methodology is in fact very successful in that it is capable of describing the local behaviors of CNT deformations. However, nonlocal formulation is indispensable beyond this local formulation for accurate descriptions of the detailed atomic behaviors involved in nanoscale phenomena such as bond breakage in fractures or in the Stone-Wales transformation. In this respect, an efficient fully nonlocal QC formulation is sought that provides a resolution down to the atomic scale as well as effective coarse-graining over local regions. Another motivation for this nonlocal QC is that it is fairly straightforward, as will be shown later. Indeed, it is free from a complex coarse-graining process at the constitutive level as well as the deformation kinematics involving exponential mapping and strain measures such as extensions and curvatures.

15.2.2. Bravais Multilattice and Inner Displacement

A graphene sheet is comprised of Bravais multilattice or composite lattice, which is formed by two simple triangular sublattices. It is specified by two base vectors $\mathbf{B_1}$ and $\mathbf{B_2}$ plus a shift vector $\mathbf{T}$, where $\mathbf{B_1}$ and $\mathbf{B_2}$ constitute a pair of base vectors on one sublattice, for example sublattice 1. The shift vector $\mathbf{T}$ indicates the distance vector from this sublattice to the other sublattice or sublattice 2 (see Figure 15-2). Let ${}_j\tilde{\mathbf{Y}}$ indicate the position vector of atom "j" measured from the reference atom site belonging to sublattice 1 on plane $\tilde{\Omega}$ in the graphene configuration $\tilde{\kappa}$ (see Figure 15-1). Here, the atom sites in a vector and tensor or in the coordinates are indicated by left lower indices, and the right left indices are reserved for their components. ${}_j\tilde{\mathbf{Y}}$ is then expressed in terms of shift vector $\mathbf{T}$ in addition to integer multiples of the base vectors $\mathbf{B_1}$ and $\mathbf{B_2}$, as follows:

$${}_i\tilde{\mathbf{Y}} = m\mathbf{B}_1 + n\mathbf{B}_2 + (\delta_i - 1)\mathbf{T} \qquad \text{with } \delta_i = 1 \text{ or } 2 \tag{15-4}$$

Here, m and n are integers. In addition, δ_i is equal to 1 for an atom belonging to sublattice 1 and to 2 for an atom belonging to sublattice 2.

For the present composite lattice, the relative displacement between the two sublattices under macroscopically uniform deformation is not specified by uniform or homogeneous strain but there is an additional internal mode of deformation. As described in Figure 15-2, the change ζ in the shift vector $\mathbf{T}$ due to deformation

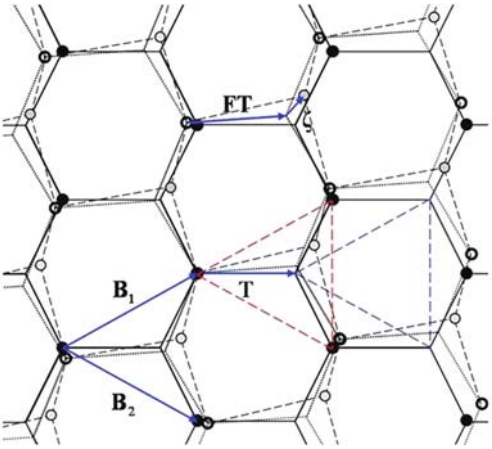


Figure 15-2. Graphene multilattice and the inner displacement between the two sublattices

of a crystalline body is termed the inner displacement between the two sublattices [13, 14]. The additional internal degrees of freedom taking place from the inner displacement field should be accounted for in energy minimization to determine an equilibrium configuration. Following Tadmor et al. [15], Park and Im [6] considered the transformation of the inner displacement ζ to $\tilde{\Omega}$ in the graphene configuration $\tilde{\kappa}$ and defined a new inner displacement field $\boldsymbol{\eta}$ referred to $\tilde{\kappa}$:

$$\boldsymbol{\eta} = \tilde{\mathbf{F}}^{-1}\boldsymbol{\zeta} \tag{15-5}$$

Here, a new variable $\boldsymbol{\eta}$ is chosen to represent the inner displacement field. The atom position vector ${}_i\tilde{\mathbf{Y}}$on the plane $\tilde{\Omega}$ in the graphene state $\tilde{\kappa}$ is then adjusted to a new vector ${}_i\mathbf{Y}$ as follows, accounting for the inner displacement $\boldsymbol{\eta}$:

$${}_i\mathbf{Y} = {}_i\tilde{\mathbf{Y}} + (\delta_i - 1)\boldsymbol{\eta} \tag{15-6}$$

Let $\bar{\Omega}_o$ and $\bar{\Omega}$ denote the tangent planes in the undeformed CNT configuration κ_o and in the deformed configuration $\kappa(t)$, respectively. Then a material line element $\Delta_{ij}\mathbf{Y} = {}_i\tilde{\mathbf{Y}} - {}_j\tilde{\mathbf{Y}} + (\delta_i - \delta_j)\boldsymbol{\eta}$ on $\tilde{\Omega}$ is mapped onto $\Delta_{ij}\mathbf{X} = {}_i\mathbf{X} - {}_j\mathbf{X}$ contained in $\bar{\Omega}_o$ in κ_o, and mapped onto $\Delta_{ij}\mathbf{x} = {}_i\mathbf{x} - {}_j\mathbf{x}$ lying on $\bar{\Omega}$ in $\kappa(t)$. Thus, for locally homogeneous deformation, this gives

$$\Delta_{ij}\mathbf{X} = \tilde{\mathbf{F}}_\mathbf{o}\Delta_{ij}\mathbf{Y} \text{ and } \Delta_{ij}\mathbf{x} = \mathbf{F}\,\Delta_{ij}\mathbf{X} = \tilde{\mathbf{F}}\,\Delta_{ij}\mathbf{Y} \tag{15-7}$$

where $\tilde{\mathbf{F}}_\mathbf{o}$, $\mathbf{F}$ and $\tilde{\mathbf{F}}$ are defined in Eq. (15-1) and Figure 15-1.

15.2.3. Interpolation Function

The preceding development shows that it is possible to account for the inner displacement field by adjusting the position vector in the graphene state $\tilde{\kappa}$, which was chosen for the reference configuration to describe the deformed state $\kappa(t)$.

For the atom position vector ${}_j\tilde{\mathbf{Y}}$ and the approximate inner displacement ${}_j\boldsymbol{\eta}^h$ on plane $\tilde{\Omega}$ in $\tilde{\kappa}$, interpolations are given in terms of the parental coordinates ${}_j\tilde{\boldsymbol{\xi}}$.

$$ {}_j\tilde{\mathbf{Y}} \;=\; \sum_{\beta=1}^{N_R} H_\beta({}_j\tilde{\boldsymbol{\xi}})\, {}_\beta\tilde{\mathbf{Y}}^R \text{ and } {}_j\boldsymbol{\eta}^h \;=\; \sum_{\beta=1}^{N_R} H_\beta({}_j\tilde{\boldsymbol{\xi}})\, {}_\beta\boldsymbol{\eta}^R \tag{15-8} $$

Here, ${}_\beta\tilde{\mathbf{Y}}^R$ and ${}_\beta\boldsymbol{\eta}^R$ indicate the position vector and the inner displacement of representative node β; furthermore, $H_\beta({}_k\tilde{\boldsymbol{\xi}})$ denotes the shape function associated with node β, and N_R and ${}_k\tilde{\boldsymbol{\xi}}$ refer to the number of representative nodes and the location or the coordinates of atom k in the standard parental domain of a finite element (see Figure 15-3), respectively. Adding these two equations results in the following equation for the adjusted position vector ${}_j\mathbf{Y}$:

$$ {}_j\mathbf{Y} \;=\; \sum_{\beta=1}^{N_R} H_\beta({}_j\tilde{\boldsymbol{\xi}})\, [{}_\beta\tilde{\mathbf{Y}}^R + (\delta_j - 1)\, {}_\beta\boldsymbol{\eta}^R] \tag{15-9} $$

In the meanwhile, for the sake of convenience in the later development, we consider another development as below by introducing a new parental coordinate ${}_j\boldsymbol{\xi}$.

$$ {}_j\mathbf{Y} \;=\; \sum_{\beta=1}^{N_R} H_\beta({}_j\boldsymbol{\xi})\, {}_\beta\tilde{\mathbf{Y}}^R \tag{15-10} $$

where ${}_j\boldsymbol{\xi}$ corresponds to the adjusted position vector ${}_j\mathbf{Y}$

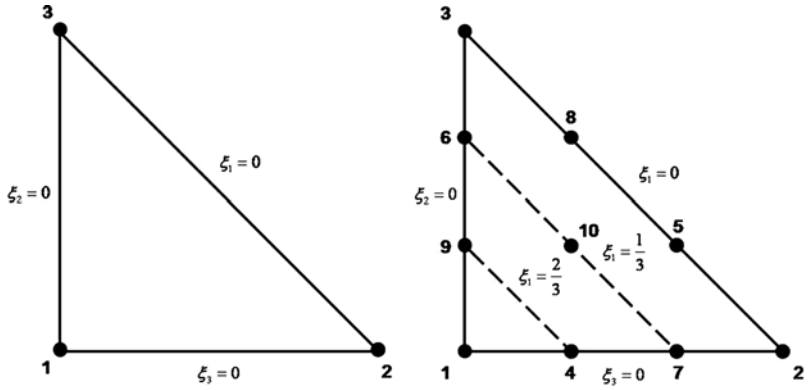


Figure 15-3. Three-noded finite element and ten-noded triangular element with triangular coordinates (ξ_1, ξ_2, ξ_3)

The link between ${}_j\tilde{\boldsymbol{\xi}}$ and ${}_j\boldsymbol{\xi}$ becomes apparent when it is noted that ${}_j\tilde{\boldsymbol{\xi}}$ may be adjusted or changed according to the inner-displacement to yield ${}_j\boldsymbol{\xi}$, just as ${}_j\tilde{\mathbf{Y}}$ is adjusted or moved to yield ${}_j\mathbf{Y}$. It should be noted that the choice of ${}_j\boldsymbol{\xi}$ instead of ${}_j\tilde{\boldsymbol{\xi}}$ in the interpolation amounts to taking into consideration the inner displacement on the parental domain.

The deformations $\tilde{\boldsymbol{\Phi}}_o$ and $\tilde{\boldsymbol{\Phi}}$ in Eq. (15-1a,b) are now reconsidered in terms of the interpolation function. To account for the inner displacement field in the graphene configuration $\tilde{\kappa}$, the deformed position vector should be given in terms of the adjusted position vector ${}_i\mathbf{Y}$, which is related to ${}_i\tilde{\mathbf{Y}}$ by Eq. (15-6). This implies that the adjusted parental coordinates ${}_j\boldsymbol{\xi}$ be employed to represent the deformed states. For isoparametric formulation, the same shape function used in Eq. (15-10) is used in terms of the adjusted parental coordinate ${}_j\boldsymbol{\xi}$:

$$ {}_k\mathbf{X}^{\mathbf{h}} = \sum_{\beta=1}^{N_R} H_\beta({}_k\boldsymbol{\xi})\, {}_\beta\mathbf{X}^{\mathbf{R}} \text{ and } {}_k\mathbf{x}^{\mathbf{h}} = \sum_{\beta=1}^{N_R} H_\beta({}_k\boldsymbol{\xi})\, {}_\beta\mathbf{x}^{\mathbf{R}} \tag{15-11} $$

Here, ${}_k\mathbf{X}^{\mathbf{h}}$ and ${}_k\mathbf{x}^{\mathbf{h}}$ denote the approximate atomic sites in κ_o and in $\kappa(t)$, respectively, and ${}_\beta\mathbf{X}^{\mathbf{R}}$ and ${}_\beta\mathbf{x}^{\mathbf{R}}$ are the undeformed and the deformed location of representative node β. This interpolation leads to the following for the displacement vector ${}_k\mathbf{u}^h$:

$$ {}_k\mathbf{u}^h = {}_k\mathbf{x}^h - {}_k\mathbf{X}^h = \sum_{\beta=1}^{N_R} H_\beta({}_k\boldsymbol{\xi})\, {}_\beta\mathbf{u}^R \tag{15-12} $$

where

$$ {}_\beta\mathbf{u}^R = {}_\beta\mathbf{x}^R - {}_\beta\mathbf{X}^R $$

The shape function $H_\beta(\boldsymbol{\xi})$ should be capable of depicting curved geometry, which is not properly handled by the linear finite element shape function employed in most QC methods applied for rectilinear crystalline materials. For finite element shape function of higher order, in this context, considered is the complete cubic polynomial in the parental coordinates. The complete cubic polynomial over the two-dimensional parental domain requires ten free constants. A triangular element with ten nodes is consistent with this interpolation. For convenience, the area coordinates or triangular coordinates [16], as shown in Figure 15-3, are utilized for routine finite element implementation. The shape functions of the 10-noded triangular element in Figure 15-3 are given in Appendix A for clarity.

If the shape function $H_\beta(\boldsymbol{\xi})$ were a complete linear function, as in the existing QC methods for applications to rectilinear crystalline materials, it would be possible to make every representative node coincide with its atom location without any difficulty. However, this may not be the case for higher order shape functions, including the cubic polynomials. For higher order elements, the best choice for the locations

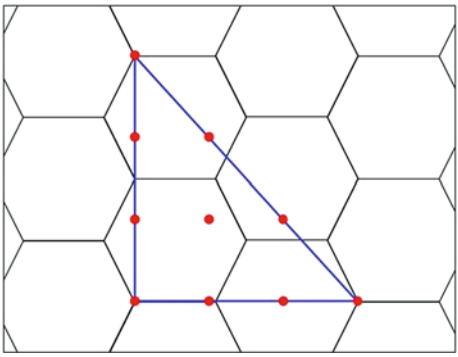


Figure 15-4. An example of atomless nodes for a ten-noded triangular element

of the physical node to be mated with the mid nodes on the parental domain is that which minimizes, over the element domain under consideration, the variation of the Jacobian determinant of the transformation between the parental domain and the physical domain.

This variation would sharply increase when the mapping between the parental and the physical domains results in a distorted physical element due to the "enforcement" of the mid nodes to be mapped onto "unnatural" atom sites. In this context, the most natural location of the interior node may not coincide with an atom site, as shown in Figure 15-4. To resolve this difficulty, Park and Im [6] introduced atomless nodal points for the cluster-based summation rule of Knap and Ortiz [4]. It is noteworthy that the energy of the cluster associated with a given representative node is well defined regardless of whether it is an atom node or atomless node. The cluster may be contemplated as a representative crystallite that possesses its energy; this is the case even for clusters with atomless nodes.

Another point concerning the atomless nodes is "Are the atomless nodes consistent with the fully nonlocal formulation?". In the local zone, CNTs behave as a structural shell. Each atom on a CNT may be thought of as being embedded onto the structural element to which it belongs. Therefore, the introduction of atomless nodes on an element subdomain is straightforward. As far as the subdivision of the nonlocal zone proceeds, the errors due to the atomless nodes continue to decrease until the limit of the fully atomistic model is reached. In the present fully nonlocal formulation, therefore, the atomless nodes do not cause any problems in the nonlocal zone as well.

15.2.4. Summation and Minimization of Energy

We assume that the system energy is represented by the summation of the node energies – the node-based summation rule [1], and by the summation of the cluster energies – the cluster-based summation rule [4]. First, for the node-based summation

rule, the following applies:

$$E^h_{\text{total}} = \sum_{\alpha=1}^{N_R} w_\alpha E_\alpha \tag{15-13a}$$

Here, w_α is the weight function for the node-based summation rule, given as

$$w_\alpha = \sum_{j=1}^{N} H_\alpha({}_j\boldsymbol{\xi}) \tag{15-13b}$$

Again N denotes the number of all atoms, which is usually much greater than the number of representative nodes N_R. Equation (15-13a) is obtained from the assumption that the energy distribution is expressed by way of interpolation:

$$E^h(\boldsymbol{\xi}) = \sum_{\alpha=1}^{N_R} H_\alpha(\boldsymbol{\xi})\, E_\alpha \tag{15-14}$$

Next, for the cluster-based summation rule, the energy is given as

$$E^h_{\text{total}} = \sum_{\alpha=1}^{N_R} (w_\alpha)_{cl}\, (E_\alpha)_{cl} \tag{15-15a}$$

where $(w_\alpha)_{cl}$ is the weight function for cluster "α", and $(E_\alpha)_{cl}$ is the cluster energy given as follows:

$$(E_\alpha)_{cl} = \sum_{j \,\in\, \text{cluster}\,\alpha} E_j \qquad \textit{with} \quad E_j = \sum_{\substack{k \\ r_{jk} < r_{cut}}} E_j(r_{jk}) \tag{15-15b}$$

Here, $E_j(r_{jk})$ represents the energy from the interaction of atom j with a neighborhood atom k, which is located within the cut-off distance of r_{cut} from atom j. For three body potentials, this may be defined similarly. Equating Eqs. (15-15a) to (15-13a), and utilizing Eqs. (15-14) and (15-15b), the following equation for the cluster weight function $(w_\alpha)_{cl}$ is obtained:

$$\sum_{\alpha=1}^{N_R} (w_\alpha)_{cl} \sum_{j \in \,\text{cluster}\,\alpha} H_\beta({}_j\boldsymbol{\xi}) = w_\beta \tag{15-15c}$$

The solution to this system of N_R equations yields the cluster weight function $(w_\alpha)_{cl}$. In the solution procedure, an efficient approximation may be made by replacing the coefficient matrix with a diagonal matrix by way of a "lumping process" based on the row-sum technique. This replaces the diagonal entries by summation of all entries on the corresponding rows, leaving all off-diagonal entries zero (see

p. 444 of [16]). The decoupling due to this diagonalization renders the solution of this system of linear equations trivial. Note that this technique is widely used for lumping mass matrices in dynamic finite element equations.

For the convergence of solution, the coarse-graining mesh is supposed to be able to represent a state of a uniform energy, which is an analog of a constant strain state in finite element method. Equation (15-14) yields the following condition, called "the partition of unity," when applied to a state of a uniform energy:

$$\sum_{\alpha=1}^{N_R} H_\alpha(\boldsymbol{\xi}) \;=\; 1 \text{ for every } \boldsymbol{\xi} \tag{15-16}$$

In addition, the following linear consistency is required:

$$\sum_{\alpha=1}^{N_R} H_\alpha(\boldsymbol{\xi})\; {}_\alpha X_j^R \;=\; X_j \tag{15-17}$$

The equilibrium configuration at the zero Kelvin temperature is obtained through the minimization of the system potential energy, which is composed of the atomic potential energy E_{total} plus the loading potential Θ:

$$\begin{aligned} \pi \;&=\; \pi({}_1\mathbf{x}, {}_2\mathbf{x}, {}_3\mathbf{x}, \ldots\ldots {}_N\mathbf{x}) \;=\; E_{\text{total}}({}_1\mathbf{x}, {}_2\mathbf{x}, {}_3\mathbf{x}, \ldots\ldots {}_N\mathbf{x}) \;+\; \Theta\,({}_1\mathbf{x}, {}_2\mathbf{x}, {}_3\mathbf{x}, \ldots\ldots {}_{N_B}\mathbf{x}) \\ &\text{with} \quad \Theta \;=\; -\sum_{j=1}^{N_B} {}_j\bar{\mathbf{f}} \cdot {}_j\mathbf{u} \end{aligned} \tag{15-18}$$

where N and N_B indicate the numbers of total atoms in the system and on the traction boundary, respectively and ${}_j\bar{\mathbf{f}}$ is the applied force on the traction boundary atom.

Introducing the QC discretization of the preceding section, the approximate potential energy can be written as a function of the current nodal vector ${}_\beta\mathbf{x}^R$ and the nodal inner displacement ${}_\beta\boldsymbol{\eta}^R$ ($\beta \;=\; 1, 2, 3 \ldots . N_R$):

$$\pi^h(\mathbf{x}^R; \boldsymbol{\eta}^R) \;=\; E^h_{\text{total}} \;+\; \Theta^h \tag{15-19}$$

In this equation, $\pi^h(\mathbf{x}^R; \boldsymbol{\eta}^R)$ is short for $\pi^h({}_1\mathbf{x}^R, {}_2\mathbf{x}^R, {}_3\mathbf{x}^R, \ldots . {}_{N_R}\mathbf{x}^R; {}_1\boldsymbol{\eta}^R, {}_2\boldsymbol{\eta}^R, {}_3\boldsymbol{\eta}^R, \ldots . {}_{N_R}\boldsymbol{\eta}^R)$. In addition, E^h_{total} is given by Eq. (15-15a), and Θ^h may be given in a similar manner. Minimization has to be taken for $6N_R$ degrees of freedom (${}_\beta x_j^R$ and ${}_\beta \eta_j^R$; *j=1, 2, 3*). For efficient computation, an iterative search is adopted such that the variation of the potential energy is made to vanish first with respect to an arbitrary variation of $\boldsymbol{\eta}^R$ while $\mathbf{x}^R$ is held constant, and is subsequently made to vanish with respect to an arbitrary variation of $\mathbf{x}^R$ while $\boldsymbol{\eta}^R$ is held constant. This alternating iterative process is repeated until solution convergence is reached. The potential energy retains its highly nonlinear behavior in terms of $\mathbf{x}^R$ and $\boldsymbol{\eta}^R$; therefore, an incremental solution procedure has to be taken to find a solution for a prescribed

loading or displacement. The undeformed state κ_o is first obtained through relaxation from the graphene state $\tilde{\kappa}$. The entire loading is then divided into many loading steps, or small incremental loadings, and the loading is then applied incrementally from the first step to the last step. In this process, the initial guess for the solution in the next step is given from the solution of the current step. For this minimization computation, limited memory BFGS (LBFGS) [17, 18], which is known as an efficient quasi-Newton method, is used.

For clarity, the process of minimization is discussed with respect to the nodal inner displacement $\boldsymbol{\eta}^R$ with $\mathbf{x}^R$ being held constant, focusing on the atomic potential energy E_{total} in the case of a two-body potential. Let $r_{ij} = \Delta_{ij}x$ indicate the length between the two atoms i and j:

$$r_{ij} = \Delta_{ij}x = \left|{}_i\mathbf{x} - {}_j\mathbf{x}\right| \tag{15-20}$$

For a two-body potential, the total atomic potential energy E^h_{total} is now expressed in terms of r_{ij}, and the minimization with respect to $\boldsymbol{\eta}^R$ is now written as:

$$\left.\frac{\partial E^h_{\text{total}}}{\partial_\beta \eta^R_k}\right|_{\hat{\mathbf{x}}^R} = \sum_{\alpha=1}^{N_R} (w_\alpha)_{cl} \frac{\partial (E_\alpha)_{cl}}{\partial r_{ij}} \left.\frac{\partial r_{ij}}{\partial_\beta \eta^R_k}\right|_{\hat{\mathbf{x}}^R} \tag{15-21}$$

where $\bullet|_{\hat{\mathbf{x}}^R}$ indicates that the nodal position vector ${}_\alpha\mathbf{x}^R$ $(\alpha = 1, 2, 3, ...N_R)$ is held constant as ${}_\alpha\mathbf{x}^R = {}_\alpha\hat{\mathbf{x}}^R$. With the aid of Eqs. (15-11) and (15-20), $\partial r_{ij}/\partial_\beta \eta^R_k\big|_{{}_\alpha\hat{\mathbf{x}}^R}$ is given in terms of the inner-displacement adjusted parental coordinate ${}_m\boldsymbol{\xi}$ as

$$\frac{\partial r_{ij}}{\partial_\beta \eta^R_k} = \sum_{\gamma=1}^{N_R} \left[\frac{\partial H_\gamma({}_i\boldsymbol{\xi})}{\partial_\beta \eta^R_k} {}_\gamma\mathbf{x}^R - \frac{\partial H_\gamma({}_j\boldsymbol{\xi})}{\partial_\beta \eta^R_k} {}_\gamma\mathbf{x}^R \right] \cdot \frac{\left({}_i\mathbf{x} - {}_j\mathbf{x}\right)}{r_{ij}} \tag{15-22}$$

Here, note that $\partial H_\gamma({}_i\boldsymbol{\xi})/\partial_\beta \eta^R_k$ is efficiently calculated, as ${}_i\boldsymbol{\xi}$ corresponds to ${}_i\mathbf{Y}$, which is adjusted by Eq. (15-9) to account for the inner displacement. That is, the high-order shape function $H_\alpha(\xi)$ is made to be consistent with the shape function of the three-node linear triangular element. To do this, the mid nodes are placed on the one-third and two-thirds points on each of the edges while the center node is on the centroid of each triangular element on the graphene plane. Thus, ${}_i\boldsymbol{\xi}$ corresponding to ${}_i\mathbf{Y}$ is picked up directly, and $H_\alpha(\xi)$ can be considered as a function of ${}_i\mathbf{Y}$:

$$H_\alpha(\boldsymbol{\xi}({}_i\mathbf{Y})) = H_\alpha({}_i\mathbf{Y}) \tag{15-23}$$

At this point, $\partial H_\gamma / \partial_\beta \eta^R_k$ is calculated through the use of Eq. (15-9). In practice, it is important to keep track of the position of each atom within each cluster to determine if any of the atoms near the boundary of an element moves to the neighboring element after its position is adjusted according to Eq. (15-9). In addition, the use of one uniform vector variable for the inner displacement field over each element,

instead of the interpolation by Eq. (15-8), is sufficient for most of the numerical examples considered.

15.2.5. Adaptive Meshing Scheme

The best use of the present QC method is made when it is combined with automatic adaptive mesh refinement. In the adaptive refinement method, the mesh keeps track of the deformation magnitude and its gradient and is increasingly refined according to the severity of the deformation and its gradient. Without this process, the mesh must be constructed based only on a priori approximate judgment regarding the overall deformation behavior. The implementation of a computational model for such automatic adaptation of a mesh demands a mesh generator for triangulation. The code TRIANGLE was employed for two-dimensional triangulation [19], which provides mesh generation with the constraint of the vertex, i.e., the nodal points, lying on the graphene domain. To estimate the deformations, the deformation measure ϵ was chosen, as follows [4]:

$$\varepsilon = \sqrt{\mathrm{II}(\mathbf{E})}h/a \tag{15-24}$$

Here, II, $\mathbf{E}$, h and a denote the second invariant of the strain, the Green strain, the element size and the bond length or the lattice spacing in the graphene, respectively. The strain $\mathbf{E}$ is simply obtainable according to its definition when the deformation gradient $\mathbf{F}$ defined in Figure 15-1 is determined (see Appendix A for a more detailed expression of the strain). If an element has a value of ϵ greater than a given tolerance, the element is refined with nodes being added on the largest edge. However, it is necessary to obtain a mesh adapted for the gradient of deformation as well as for the magnitude of deformation so that a steep gradient of strains may be properly captured in the numerical solution. This is consistent with the basic premise for the coarse-grained local region, in which the deformation is free from any severe abrupt changes. In addition, the gradient of deformation in the element sub-domain is not constant, unlike the linear triangular element, due to the high-order element employed for the interpolation of the displacement field. The gradient $\nabla\mathrm{II}$ of the second invariant II at the Gauss points of the high-order element is suitable as a measure of the gradient of deformation. If the measure $\nabla\mathrm{II}$ in the sub-domain of an element is larger than a given tolerance, the element is subjected to refinement.

15.2.6. Deformation of Multiwalled Carbon Nanotubes (MWCNTs)

The formulation discussed thus far may be applied for the coarse-graining of MWCNTs. Rather than introducing any solid elements, the modeling of each layer with the preceding ten-node element is preferred. To reduce the total number of degrees of freedom, it is advantageous to have the node distribution lined up along the radial direction so that, along the radial line from the tube center to a node on the outer most layer, each layer has a nodal point wherever it meets with the radial line.

In this way, Park and Im [6] conducted quasicontinuum simulations of the bending and the torsion of a 15-walled CNT with 673,650 atoms or with over 2×10^6 degrees of freedom. The QC model for this simulation has only 291,244 degrees of freedom, which is less than 13% of the fully atomistic model. The rippling structures reported in experiments [20] were well captures (see [6] for detail).

15.2.7. Numerical Examples

In this section, several numerical examples are shown to demonstrate the accuracy and effectiveness of the present QC scheme. First, the force field employed for the computation is briefly reviewed. A case of simple bending deformation is then considered, in which comparison is made with the results from the fully atomistic simulation or the molecular mechanics simulation. A simply supported CNT pressed by an AFM tip at the center is then chosen as an example of the QC combined with the adaptive refinement method discussed in Section 15.2.5.

15.2.7.1. Bonding and Nonbonding Interaction for CNT

The Tersoff-Brenner potential has been widely used to study the mechanics of carbon nanotubes and other low-dimensional carbon nanostructures. As a bond-order potential, it depends not only on the interatomic distances but also on interatomic bond angles. The potential was first suggested by Tersoff for covalent systems such as carbon (C), silicon (Si) and germanium (Ge) [21] and was then elaborately modified by Brenner for hydrocarbons [22]. The Tersoff-Brenner potential can be written compactly as

$$E_{\text{Bond}} = \sum_i \sum_{j(>i)} \left[V_R(r_{ij}) - \bar{B}_{ij} V_A(r_{ij}) \right] \tag{15-25}$$

where r_{ij} denotes the distance between bond connecting atoms i and j. Furthermore, V_R and V_A are the repulsive and attractive pair terms, respectively, and are given as

$$V_R(r_{ij}) = f_c(r_{ij}) \frac{D_e}{S-1} e^{-\beta\sqrt{2S}(r_{ij}-r_e)} \tag{15-26a}$$

$$V_A(r_{ij}) = f_c(r_{ij}) \frac{SD_e}{S-1} e^{-\beta\sqrt{2/S}(r_{ij}-r_e)} \tag{15-26b}$$

$$\bar{B}_{ij} = (B_{ij} + B_{ji})/2 + F_{ij}(N_i^{(t)}, N_j^{(t)}, N_{ij}^{\text{conj}}) \tag{15-26c}$$

where the parameters D_e, S, β and r_e are determined from the known physical properties of hydrocarbons. The function f_c is a smooth cut-off function that limits the range of interaction of atoms and bond breaking. This function smoothly goes to zero between $R^{(1)} = 0.17$ nm and $R^{(2)} = 0.2$ nm. The function $\bar{B}_{ij}$ represents a multi-body coupling for the bonding. These parameters and functions are presented in Appendix B. The nonbonding interaction between tube walls plays an important role in the mechanical behavior of multi-walled carbon

nanotubes. In this study, the Lennard-Jones type potential is used for nonbonding interaction [23].

$$E_{VDW} = \sum_i \sum_{j(>i)} 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{15-27}$$

Where, ε and σ are parameters related to the energy and length according to the materials. The nonbonding parameters for carbon nanotubes are also given in Appendix B.

15.2.7.2. Bending Simulations for a SWCNT

The first numerical example using the generalized quasicontinuum method is the bending simulation for a SWCNT. A tube 43.36 nm long with chirality (60,0) is modeled for the bending simulation. The tube is bent by imposing displacement on each end. The loading conditions used in this paper are not pure bending. In a pure bending case, a 180° bent tube forms a half circle. In this study, however, the tube is subjected to bending plus compression such that the two tangent lines at the both ends meet each other at the center with each of the tangent line segment having half the CNT length (see Figure 15-5). The maximum angle of bending in this simulation is 25°. This example was employed by Park and Im [6], and it demonstrates the adequacy of the proposed method.

The coarse-grained model for bending simulation has a total number of degrees of freedom of 30,360 (including the inner displacement of 1,560), while the full atomistic or MM model has 72,000 degrees of freedom. Figure 15-6 shows that the strain energy from the QC is in good agreement with that from molecular mechanics (MM). When the bending angle reaches 10°, the strain energy in Figure 15-6 is released and a kink occurs at the center. The relative difference in the strain energy between the MM and the QC just before and directly after kinking is 1.48 and 3.96%, respectively. Figure 15-7 shows the equilibrium configurations of atoms and elements at bending angles of 0°, 9.5°, 10° and 25°. Although the inner displacement in the coarse-graining model for CNTs occupies a very small number

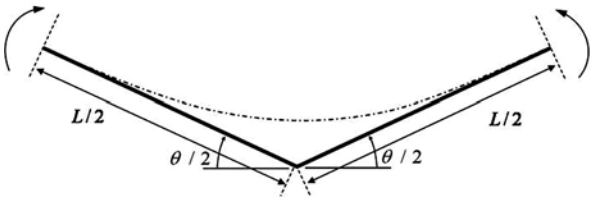


Figure 15-5. Bending plus compression of a CNT under consideration (L=the initial length of the CNT): The *dashed-dotted line* indicates the centroidal line of the tube in the deformed configuration (Reprinted with permission from Park and Im [6]. Copyright (2008) by the American Physical Society. http://link.aps.org/abstract/PRB/v77/p184109)

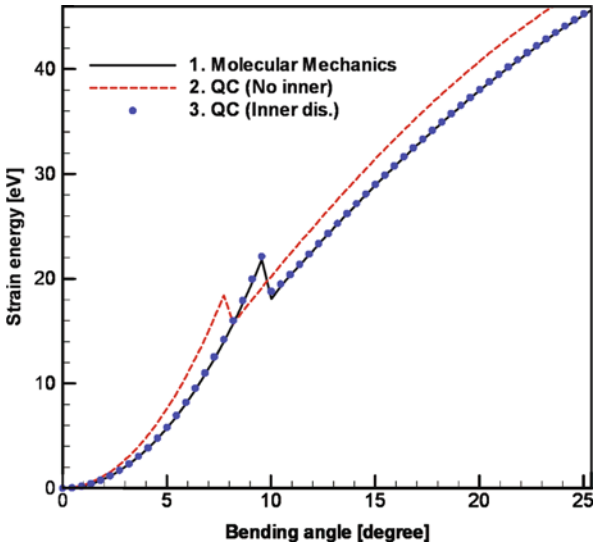


Figure 15-6. Strain energy curve with respect to bending angle for (60,0) SWCNT. The *solid black line* indicates the result using molecular mechanics, and the *dotted red line* quasicontinuum without considering the inner displacement. The *blue circle* denotes quasicontinuum result with the inner displacement

of degrees of freedom, it has a substantial effect on the deformation and on the strain energy curve (See Figure 15-6). In this and all subsequent examples, the inner displacement degrees of freedom $\boldsymbol{\eta}$ are assumed to be uniform throughout each individual elements; this approximation was found to incur only a negligibly small error compared with the case in which $\boldsymbol{\eta}$ is interpolated in terms of the nodal degrees of freedom by the same high-order function used for the deformed position vector $\mathbf{x}$. Moreover, only one iteration for $\mathbf{x}^R$ and $\boldsymbol{\eta}^R$ was carried out to determine the equilibrium configuration through minimization. The result obtained after this single iteration is indistinguishable from that of a converged solution. In Figure 15-6, the solid black line indicates the result using molecular mechanics, and the dotted red line quasicontinuum without considering the inner displacement. The blue circle denotes quasicontinuum result with the inner displacement.

Recently, the electromechanical properties of CNTs have attracted much interest, largely based on their potential for application to ultra-sensitive electromechanical sensors. Among others, Tombler et al. [24] demonstrated that the electrical conductance of a CNT changes when the center part of a suspended nanotube is deformed by a sharp AFM tip. To simulate the mechanical behavior of a CNT under this circumstance, a simply-supported CNT subjected to indentation by AFM tip is considered, as shown in Figure 15-8. The CNT under consideration is of a single-walled armchair type with its chirality being (20,20). It is comprised of 15,960 carbon atoms. The radius is approximately 3 nm and the length is 50 nm.

Both ends are placed on the support, and the center is gradually indented with the AFM tip, as represented by a repulsive potential, up to the maximum indentation depth of 3.5 nm. Figure 15-8 shows the adaptive refinement taking place according

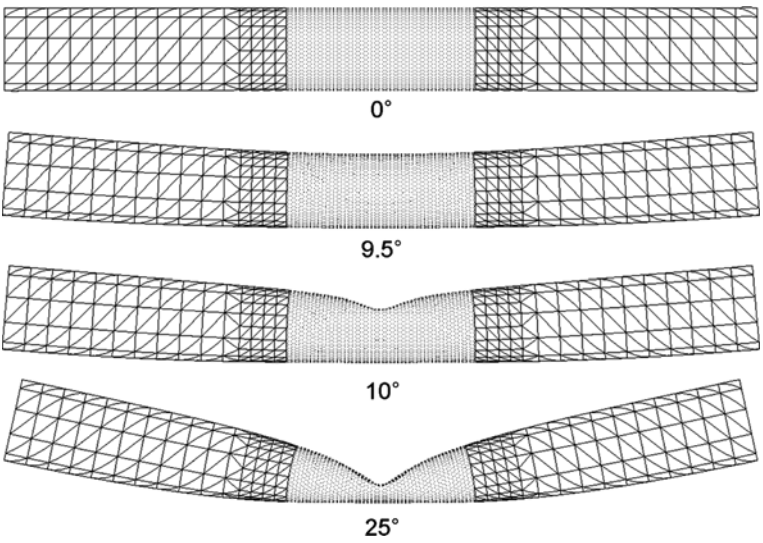


Figure 15-7. Equilibrium configurations of atoms and meshes using a quasicontinuum model with inner the displacement at each bending angle

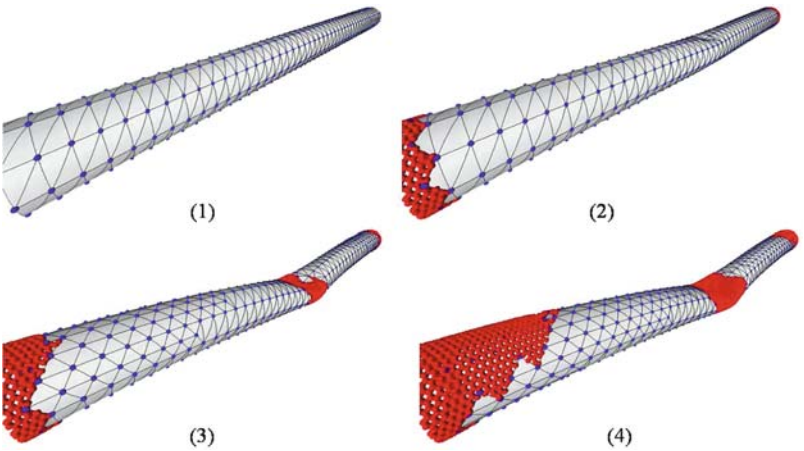


Figure 15-8. The (20,20) SWCNT result of simulation for AFM indentation employing adaptive meshing refinement

to the increasing indentation depth. The initial mesh starts with a total number of degrees of freedom of 11,907. The final mesh has 24,432 degrees of freedom, which is nearly 51% of the fully atomistic model. Figure 15-9 shows a plot of the energy versus the AFM tip displacement. The figure shows that the result from the present adaptive QC is in good agreement with the molecular mechanics result.

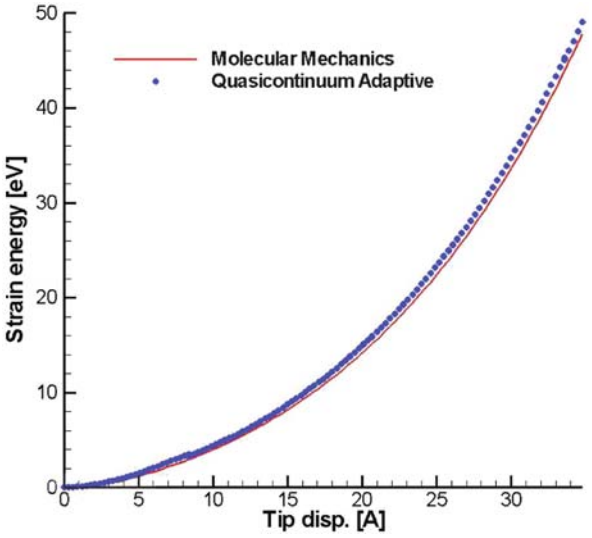


Figure 15-9. Strain energy curve with respect to the AFM tip indenting depth for the (20,20) SWCNT: The *solid red line* indicates the result using molecular mechanics, and the *blue dotted line* indicates the quasicontinuum result employing the adaptive mesh refinement

15.3. QC METHOD FOR CNTS BY USE OF VARIABLE-NODE ELEMENTS

In this section, the possibility of generalizing further the aforementioned work of Park and Im [6] for more effective computation is explored. To this end, new high order finite elements with variable nodes are introduced [7–9] and the interelement compatibility is relaxed, which will cause a negligible error with an increasing number of iterations for minimization. It is shown that the QC scheme based on the variable-node elements yields accurate solutions in comparison to solutions from molecular mechanics or from the previous QC approach by way of the triangular element. The effectiveness of the variable-node elements for the QC method is discussed below.

15.3.1. Variable Node Elements for QC

In the conventional quasicontinuum method, triangular and tetrahedron elements are employed for the coarse-graining of a two- and a three-dimensional domain, respectively. However, it is well known that quadrilateral or hexahedral elements show much better performance than triangular or tetrahedron elements in a finite element method. Nevertheless, the conventional quasicontinuum method has adopted triangular or tetrahedron elements for convenience of adaptive refinement. This adaptive refinement in the presence of quadrilateral or hexahedral elements is easily handled by variable-node elements, which have been successfully employed with regards to

non-matching problems and for bridging between two shape functions of different orders for smooth transition (see [7–9] and the references therein).

Triangular and tetrahedron elements are much less efficient compared to quadrilateral and hexahedral elements in terms of element searching. The displacement of ordinary atoms in the local region is determined via interpolation of the representative nodes under the Cauchy-Born rule. Consequently, every ordinary atom memorizes the identification number of the element to which the atom belongs. For instance, in the two-dimensional case with triangular elements, ordinary atoms must check every element in their neighborhood to find the one that contains itself. To do this, the equality of $A_{\text{total}} = A_1 + A_2 + A_3$ is checked, and the value of total A_{total} is computed by the position vectors of the three-element nodes. Each of the areas A_1, A_2, A_3, corresponding to the triangular or area coordinates ξ_1, ξ_2, ξ_3, is determined by the atom position and the associated two nodal position vectors. On the other hand, for quadrilateral elements, the element containing a given atom is easily found by comparing the position or the parental coordinates of the atom with the positions of the representative nodes in terms of the parental coordinates without computing the areas.

Despite the aforementioned advantage of quadrilateral and hexahedron elements, they are not commonly used in finite element simulations involving adaptive refinement as in the QC method, whereas triangular and tetrahedron elements are frequently employed in this case. This is due to the fact that realizing the adaptive refinement for a mesh comprised of quadrilateral or hexahedron elements is not straightforward. Recently, Lim et al. [7–9] showed that "nonmatching meshes," are efficiently treated utilizing variable-node elements. This suggests that the use of variable-node elements makes it possible to introduce a finite element mesh, as shown in Figure 15-10, which is composed of quadrilateral elements. This appears to violate the interelement compatibility according to the notion of the conventional finite elements. However, the nodal points are shared by the two neighboring elements, and their shape functions are constructed for low-order elements such that the interelement compatibility is satisfied [7–9]. Recently, Kwon et al. [25] employed three-dimensional variable-node elements in their QC for rectilinear crystalline materials. This enables one to implement the adaptive refinement for hexahedron elements in a straightforward manner, as the interelement compatibility is taken care of by the variable node elements. Furthermore, an additional mesh generation program, which is necessary for triangular and tetrahedron elements, is not required for quadrilateral and hexahedral elements, and the adaptive refinement may be continued by adding elements and nodes where refinement is needed.

Finite elements of orders higher than a linear polynomial are essential to model a curved geometry such as that of CNTs. From the successful modeling of CNTs with the 10-noded triangular element of the complete cubic polynomial in Section 15.2, a 12-noded isoparametric element, as shown in Figure 15-11, appears to be capable of modeling CNTs. Some variable node elements might be constructed such that they may take care of nonmatching meshes involving the 12-noded isoparametric elements, satisfying the interelement compatibility. In such circumstances,

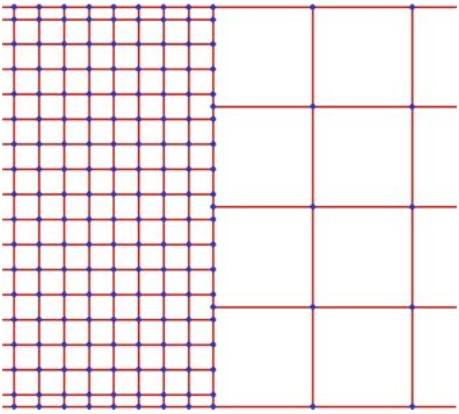


Figure 15-10. An example of an application of variable-node elements for a non-matching mesh

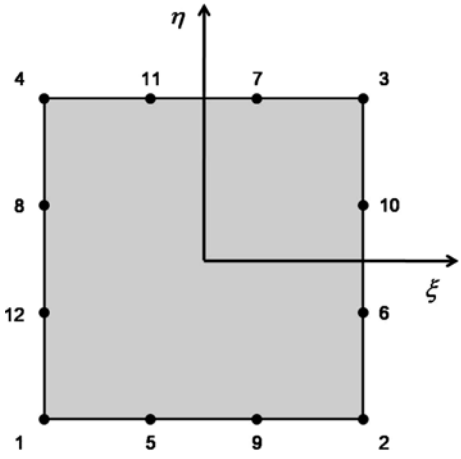


Figure 15-11. A 12-noded isoparametric element on the parental domain (ξ,η)

the issue of the adaptive refinement would be completely resolved by use of these variable node elements, as shown in Figure 15-13. However, the construction of this type of variable node elements would be costly because of the complexity of the shape functions. Here, simple variable-node elements with hierarchical structure are instead devised with 15, 18, 21 and 24 nodes, as shown in Figure 15-12. They are constructed such that their shape functions may all be given as polynomial type functions. Thanks to this straightforward shape functions, the coding is very simple, but the exact interelement compatibility is sacrificed in the course of the adaptive refinement by using these elements. That is, the interelement compatibility is neglected when it is assembled with two 12-noded elements along the edge having extra nodes, as in Figure 15-13.

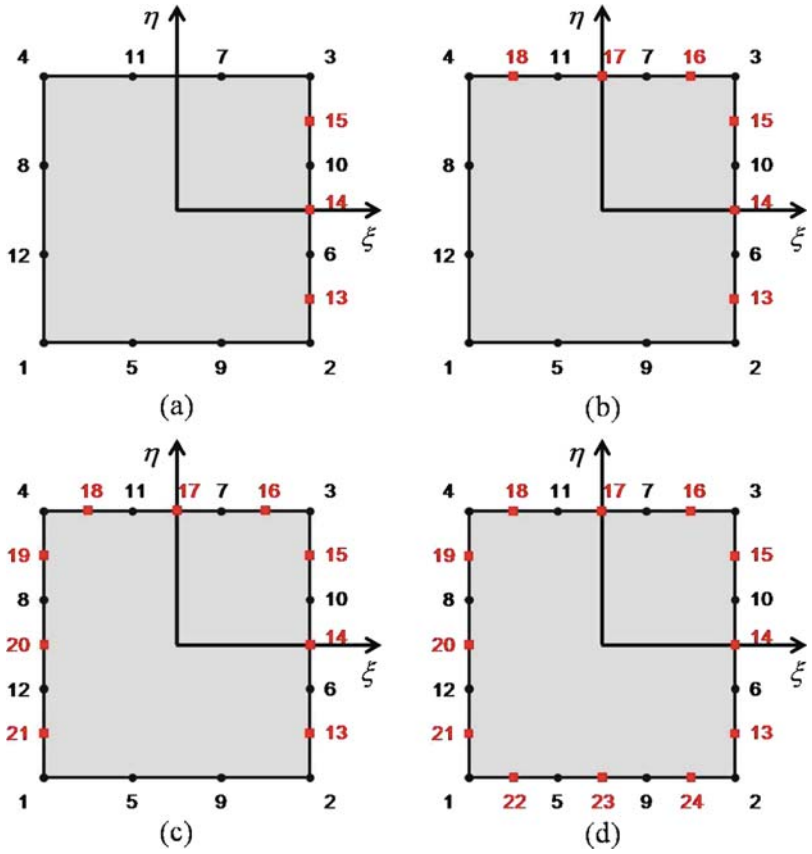


Figure 15-12. Variable-node elements for adaptive refinement on the parental domain (*squares* indicate the nodes added): (**a**) A 15-noded variable-node element, (**b**) A 18-noded variable-node element, (**c**) A 21-noded variable-node element, (**d**) A 24-noded variable-node element

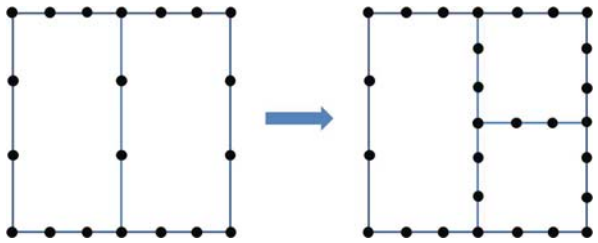


Figure 15-13. Mesh refinement by way of the variable-node element: two 12-noded elements are refined to one 15-noded variable-noded element and two 12-noded elements

The construction of the shape functions for these variable-node elements starts with the shape functions of the 12-noded element, which is given in Appendix C. Here, in addition to the partition of unity Eq. (15-16) and the linear consistency Eq. (15-17), the shape functions of the 12-noded and the variable node elements satisfy the following Kronecker delta condition:

$$H_\alpha({}_\beta\boldsymbol{\xi}) = \delta_{\alpha\beta} \tag{15-28}$$

Here, H_α and ${}_\beta\boldsymbol{\xi}$ are the shape functions associated with the α-th node and the position of the β-th node, respectively. For the variable node elements to continue to fulfill this condition, the shape functions should be modified systematically as new nodes are added, as described in Appendix C.

As aforementioned, unlike the lower order variable-node elements reported in earlier studies [7–9] and [25], the present variable-node elements fail to meet the interelement compatibility in the circumstances like Figure 15-13. However, this does not imply that they are not applicable for the present nonlocal QC scheme, in which the potential energy is calculated from the discrete atomic view and not from the view of the continuum shell. The energy is not given as a direct function of nodal displacements in the context of finite element method for continuum shells; however, it is computed by summing up the individual energies of the atoms in the discrete manner within each cluster. As a consequence of this, the potential energy is not significantly sensitive to the compatibility condition. This will be confirmed by the numerical examples in the next section, in which the solutions using the present variable-node elements are compared with those from the triangular elements and from the molecular mechanics. Once the interelement compatibility is neglected, adaptive refinement is straightforwardly implemented, as in the case of the triangular element in Section 15.2.7. New nodes are inserted where new elements are added according to the two refinement criteria of the strain magnitude and its gradient.

15.3.2. Numerical Examples

In this section, an example is considered to confirm the accuracy of the 12-noded isoparametric or rectangular element and the associated variable-node elements with regards to the QC method applied for CNT bending. Subsequently, adaptive meshing for the bending of a CNT under loading, as shown in Figure 15-5, will be adopted for a comparison between the QC method with the present rectangular elements and that with the triangular element discussed in Section 15.2.

First, for the bending of a CNT, the chirality is (24,0) and the length and the diameter is 24.2 and 1.9 nm. The loading condition is given in terms of the prescribed displacement on both ends. The total number of atoms is 5,376, corresponding to 16,128 degrees of freedom. The number of degrees of freedom for the QC model is 7,680. The undeformed mesh and several deformed configurations are shown in Figure 15-14. The energy versus the bending angle, as defined in Figure 15-5, is

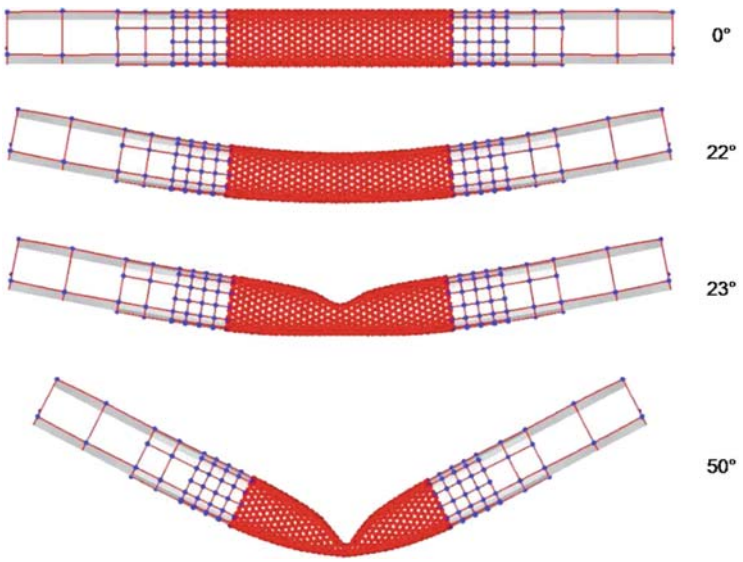


Figure 15-14. Equilibrium configurations of atoms and meshes using the quasicontinuum with variable-node elements at various bending angles

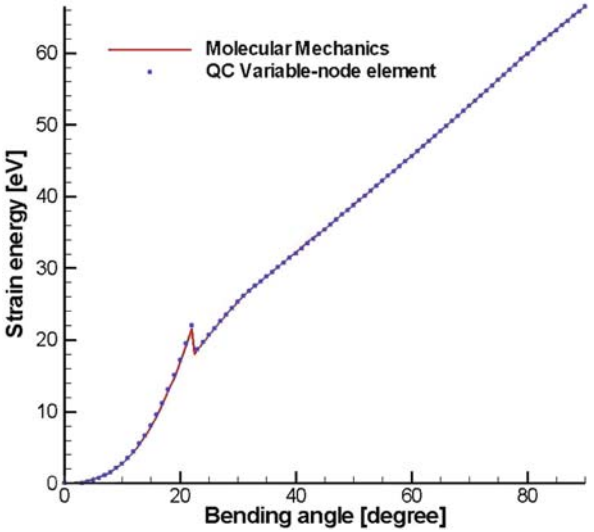


Figure 15-15. Strain energy curve with respect to the bending angle for a (24,0) SWCNT: The *solid red line* indicates the result using molecular mechanics, and the *dotted blue circle* indicates the quasicontinuum with variable-node elements

presented in Figure 15-15, in which the comparison is made between the two solutions; the first is the solution from the QC with the present rectangular elements, and the second is from molecular mechanics. The good agreement between the two solutions is apparent.

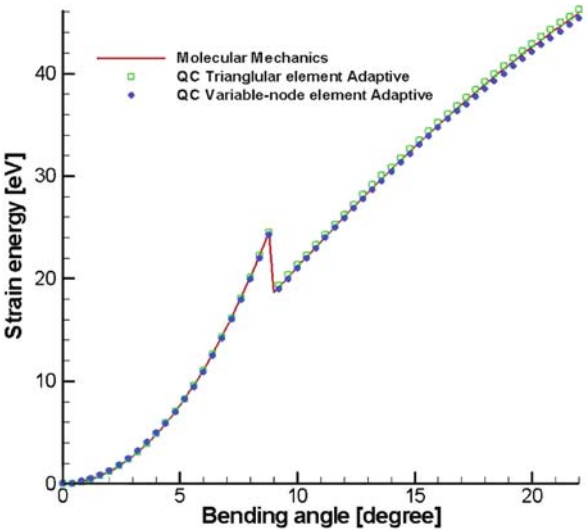


Figure 15-16. Strain energy curve with respect to bending angle for (40,40) SWCNT. The *solid red line* indicates the result using molecular mechanics, and the *green square* quasicontinuum with adaptive meshing refinement employing the triangular elements. The *blue circle* denotes the quasicontinuum with adaptive meshing refinement employing the rectangular and the variable-node elements

The next example considered is the adaptive simulation for the bending of an armchair type CNT with chirality (40,40) under the loading shown in Figure 15-5. The tube length and the diameter are 52.34 and 5.5 nm, respectively. The total number of the carbon atoms is 33,360, which corresponds to 100,080 degrees of freedom. The number of the nodes of the initial mesh is 984, corresponding to 2,952 degrees of freedom. The number of the total degrees of freedom increases to 36,510, including the inner displacement degrees of freedom, when the bending angle reaches 22° (see Figure 15-17). In actuality, this example was chosen for adaptive refinement in the QC using the triangular element by Park and Im [6]. The present solution from the QC by the 12-noded rectangular element is compared with each of the solutions from [6] and from molecular mechanics in Figure 15-16. The three solutions are shown to be in good agreement. The initial undeformed mesh and several deformed states are presented in Figure 15-17. The QC by the rectangular and the variable-node element is much more efficient in terms of the solution time than the QC by the triangular elements in Section 15.2 and in Park and Im [6]. Furthermore, a mesh generation code is unnecessary for the rectangular element case, whereas the mesh generator TRIANGLE was used for the triangular element case.

15.4. CONCLUSIONS

After a brief review on the QC method applied for the deformations of curved structures such as CNTs, this work proposes a new implementation of the QC for curved bodies using the 12-noded rectangular element and associated variable-node

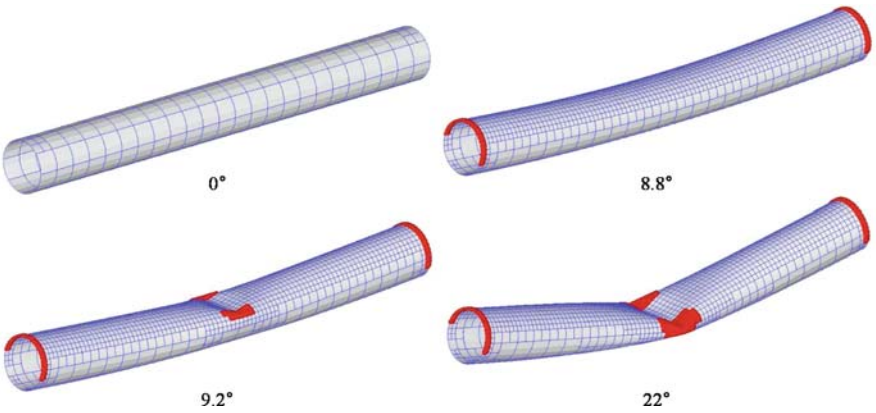


Figure 15-17. Equilibrium configurations of atoms and meshes using the quasicontinuum model with adaptive refinement using variable-node elements at each bending angle

elements. This implementation is much more efficient than the use of the triangular element, as in Park and Im [6]. Particularly, the use of these elements renders it straightforward to build an adaptively fine mesh in proportion to the severity of deformation. Numerical examples demonstrate the effectiveness and the accuracy of the proposed rectangular type elements in the QC for CNTs.

In passing, it should be pointed out that the strength of this kind of approach utilizing the variable-node elements stands out also for the QC for rectilinear crystalline materials. An example is shown in the nano-indentation of a single crystal metal substrate [25]. Complicated adaptive refinement for the three-dimensional domain of the substrate is fully automatic, and the resulting QC code is extremely easy to apply for a variety of problems.

In addition, the present QC can be combined with a DFT calculation for multiscale computing of QC/DFT hybridization to account for the effect of electronic structures, which are important in exploring fracture behaviors or electromechanical coupling phenomena (see Ref. [26] for detail).

ACKNOWLEDGMENT

This research was supported by NRL (National Research Lab) program through the Korea Science and Engineering Foundation funded by the Ministry of Education, Science and Technology (R0A-2007-000-20115-0).

APPENDIX A. THE GREEN STRAIN IN DEFORMATION OF A CNT

The strain **E** in Eq. (15-24) is obtainable if the deformation gradient **F** from tangent plane $\bar{\Omega}_o$ of the undeformed tube to tangent plane $\bar{\Omega}$ of the deformed tube (see Figure 15-1) is determined:

$$\mathbf{E} = 1/2(\mathbf{F}^{\mathrm{T}}\mathbf{F} - \mathbf{I}) \tag{A-1}$$

We denote by $\tilde{\Phi}_o$ the point mapping of parental domain $\tilde{\Omega}$ onto undeformed surface Ω_o, and by $\tilde{\Phi}$ onto deformed surface Ω. Based on Eq. (15-11), the mapping $\tilde{\Phi}$ is written as follows;

$$\mathbf{x} = x^i \mathbf{e}_i = \tilde{\Phi}^i(\boldsymbol{\xi})\mathbf{e}_i \quad \boldsymbol{\xi} = \xi^I \mathbf{E}_I \tag{A-2a}$$

$$\tilde{\Phi}^i(\boldsymbol{\xi}) = \sum_{\alpha=1}^{N_R} H_\alpha(\boldsymbol{\xi})\, {}_\alpha x_i^R \tag{A-2b}$$

where i=1,2,3 and I=1,2. The mapping $\tilde{\Phi}$ maps the point ξ, which is the inner displacement adjusted in the original graphene, onto the position $\mathbf{x}$ of the deformed tube. The mapping of $\tilde{\Phi}_o$, which builds an initial cylinder-shaped tube, is defined by

$$\mathbf{X} = X^i \mathbf{e}_i = \tilde{\Phi}_o^i(\boldsymbol{\xi})\mathbf{e}_i \tag{A-3a}$$

$$\tilde{\Phi}_o^1(\boldsymbol{\xi}) = \hat{r}\cos\theta = \hat{r}\cos(\xi^1/\hat{r}) \tag{A-3b}$$

$$\tilde{\Phi}_o^2(\boldsymbol{\xi}) = \hat{r}\sin\theta = \hat{r}\sin(\xi^1/\hat{r}) \tag{A-3c}$$

$$\tilde{\Phi}_o^3(\boldsymbol{\xi}) = \xi^2 \tag{A-3d}$$

where $\hat{r}$ is the radius of the tube, which is initially determined. In this case, the convected basis vectors which are tangent to surfaces Ω_o and Ω are

$$\mathbf{G}_I = \frac{\partial \tilde{\Phi}_o^i}{\partial \xi^I}\mathbf{e}_i \; \textit{and} \; \mathbf{g}_I = \frac{\partial \tilde{\Phi}^i}{\partial \xi^I}\mathbf{e}_i \tag{A-4}$$

For example, the tangent vectors to the undeformed tube surface Ω_o are given by

$$\mathbf{G}_1 = -\sin(\xi^1/\hat{r})\mathbf{e}_1 + \cos(\xi^1/\hat{r})\mathbf{e}_2, \quad \mathbf{G}_2 = \mathbf{e}_3, \tag{A-5}$$

The deformation gradient $\mathbf{F}$ is represented as

$$\mathbf{F} = \mathbf{g}_I \otimes \mathbf{G}^I \tag{A-6}$$

Straightforwardly, the Green strain in Eq. (A-2) is written as

$$\mathbf{E} = \frac{1}{2}\left(g_{IJ}\mathbf{G}^I \otimes \mathbf{G}^J - G_{IJ}\mathbf{G}^I \otimes \mathbf{G}^J\right) \tag{A-7}$$

Here, two convected base vectors $\mathbf{G}_I$ (I=1, 2) in Eq. (A-4) are orthonormal to each other, and $\mathbf{G}^I$ is then identical to $\mathbf{G}_I$, so that G_{IJ} reduces to the Kronecker delta.

APPENDIX B. THE FUNCTIONS AND THE PARAMETERS IN THE TERSOFF-BRENNER POTENTIAL

The cutoff function in the Tersoff-Brenner potential is given by

$$f_c(r_{ij}) = \begin{cases} 1 & r < R^{(1)} \\ \frac{1}{2}\left\{1 + \cos\left[\frac{\pi(r - R^{(1)})}{R^{(2)} - R^{(1)}}\right]\right\} & R^{(1)} < r < R^{(2)} \\ 0 & r > R^{(2)} \end{cases} \tag{B-1}$$

The function B_{ij} of Eq. (15-26c) is defined as

$$B_{ij} = \left[1 + \sum_{k(\neq i,j)} G(\theta_{ijk}) f_c(r_{ik})\right]^{-1/2} \tag{B-2}$$

where θ_{ijk} is the angle between bonds $i-j$ and $i-k$. In this chapter, B_{ij} is simplified due to the ignorance of the hydrocarbon interaction and the consideration of the solid state carbon. The function G is given by

$$G(\theta_{ijk}) = a_o\left[1 + \frac{c_o^2}{d_o^2} - \frac{c_o^2}{d_o^2 + (1 + \cos\theta)^2}\right] \tag{B-3}$$

The parameters are determined from the known physical properties of hydrocarbons and are presented as follows:

$D_e = 6.0\ eV\ \ S = 1.22\ \ \beta = 2.1\ \text{Å}^1\ r_e = 1.39\ \text{Å}$
$R^{(1)} = 1.7\ \text{Å}\ R^{(2)} = 2.0\ \text{Å}\ a_o = 0.00020813\ \ c_o = 330\ \ d_o = 3.5\ .$

The function $F_{ij}(N_i^t, N_j^t, N_{ij}^{conj})$ in Eq. (15-26c) indicates the correction term for carbon-carbon bonds (see [22] for detail), but this term has been left out of simplicity of computation in the present computation.

APPENDIX C. THE SHAPE FUNCTIONS FOR A 24-NODED VARIABLE-NODE ELEMENT

The shape functions of a 12-noded isoparametric element are known and are shown as below:

$$\phi_1^0(\xi,\eta) = \frac{1}{32}(1-\xi)\ (1-\eta)\ \left(-10 + 9\xi^2 + 9\eta^2\right)$$

$$\phi_2^0(\xi,\eta) = \frac{1}{32}(1+\xi)\ (1-\eta)\ \left(-10+9\xi^2+9\eta^2\right)$$

$$\phi_3^0(\xi,\eta) = \frac{1}{32}(1+\xi)\ (1+\eta)\ \left(-10+9\xi^2+9\eta^2\right)$$

$$\phi_4^0(\xi,\eta) = \frac{1}{32}(1-\xi)\ (1+\eta)\ \left(-10+9\xi^2+9\eta^2\right) \tag{C-1}$$

$$\phi_5^0(\xi,\eta) = \frac{9}{32}(1-3\xi)\ (1-\eta)\ \left(1-\xi^2\right)$$

$$\phi_6^0(\xi,\eta) = \frac{9}{32}(1+\xi)\ (1-3\eta)\ \left(1-\eta^2\right)$$

$$\phi_7^0(\xi,\eta) = \frac{9}{32}(1+3\xi)\ (1+\eta)\ \left(1-\xi^2\right)$$

$$\phi_8^0(\xi,\eta) = \frac{9}{32}(1-\xi)\ (1+3\eta)\ \left(1-\eta^2\right)$$

$$\phi_9^0(\xi,\eta) = \frac{9}{32}(1+3\xi)\ (1-\eta)\ \left(1-\xi^2\right)$$

$$\phi_{10}^0(\xi,\eta) = \frac{9}{32}(1+\xi)\ (1+3\eta)\ \left(1-\eta^2\right)$$

$$\phi_{11}^0(\xi,\eta) = \frac{9}{32}(1-3\xi)\ (1+\eta)\ \left(1-\xi^2\right),$$

$$\phi_{12}^0(\xi,\eta) = \frac{9}{32}(1-\xi)\ (1-3\eta)\ \left(1-\eta^2\right)$$

The shape functions for additional nodes from nodes 13 to 24 can be devised such that the partition of unity and the Kronecker delta condition may be satisfied.

$$\phi_{13}(\xi,\eta) = \frac{9}{40}(1+\xi)\ \left(1-\eta^2\right)\ \left(1-9\eta^2\right)\ \left(1-\frac{3}{2}\eta\right)\eta$$

$$\phi_{14}(\xi,\eta) = \frac{1}{2}(1+\xi)\ \left(1-\eta^2\right)\ \left(1-9\eta^2\right)\ \left(1-\frac{9}{4}\eta^2\right)$$

$$\phi_{15}(\xi,\eta) = -\frac{9}{40}(1+\xi)\ \left(1-\eta^2\right)\ \left(1-9\eta^2\right)\ \left(1+\frac{3}{2}\eta\right)\eta$$

$$\phi_{16}(\xi,\eta) = -\frac{9}{40}(1+\eta)\ \left(1-\xi^2\right)\ \left(1-9\xi^2\right)\ \left(1+\frac{3}{2}\xi\right)\xi$$

$$\phi_{17}(\xi,\eta) = \frac{1}{2}(1+\eta)\left(1-\xi^2\right)\left(1-9\xi^2\right)\left(1-\frac{9}{4}\xi^2\right)$$

$$\phi_{18}(\xi,\eta) = \frac{9}{40}(1-\eta)\left(1-\xi^2\right)\left(1-9\xi^2\right)\left(1-\frac{3}{2}\xi\right)\xi$$

$$\phi_{19}(\xi,\eta) = -\frac{9}{40}(1-\xi)\left(1-\eta^2\right)\left(1-9\eta^2\right)\left(1+\frac{3}{2}\eta\right)\eta \qquad \text{(C-2)}$$

$$\phi_{20}(\xi,\eta) = \frac{1}{2}(1-\xi)\left(1-\eta^2\right)\left(1-9\eta^2\right)\left(1-\frac{9}{4}\eta^2\right)$$

$$\phi_{21}(\xi,\eta) = \frac{9}{40}(1-\xi)\left(1-\eta^2\right)\left(1-9\eta^2\right)\left(1-\frac{3}{2}\eta\right)\eta$$

$$\phi_{22}(\xi,\eta) = \frac{9}{40}(1-\eta)\left(1-\xi^2\right)\left(1-9\xi^2\right)\left(1-\frac{3}{2}\xi\right)\xi$$

$$\phi_{23}(\xi,\eta) = \frac{1}{2}(1-\eta)\left(1-\xi^2\right)\left(1-9\xi^2\right)\left(1-\frac{9}{4}\xi^2\right)$$

$$\phi_{24}(\xi,\eta) = -\frac{9}{40}(1-\eta)\left(1-\xi^2\right)\left(1-9\xi^2\right)\left(1+\frac{3}{2}\xi\right)\xi$$

The shape functions for nodes 1–12 are now modified by subtracting the above additional shape functions multiplied by the weight from the original shape functions of a 12-noded element $\phi_j^0 (j = 1 \sim 12)$. The shape functions of a 24-noded variable-node element are finally written as in Eq. (C-3).

$$\phi_1(\xi,\eta) = \phi_1^0 - \frac{1}{16}\phi_{19} + \frac{1}{16}\phi_{20} - \frac{5}{16}\phi_{21} - \frac{5}{16}\phi_{22} + \frac{1}{16}\phi_{23} - \frac{1}{16}\phi_{24}$$

$$\phi_2(\xi,\eta) = \phi_2^0 - \frac{5}{16}\phi_{13} + \frac{1}{16}\phi_{14} - \frac{1}{16}\phi_{15} - \frac{1}{16}\phi_{22} + \frac{1}{16}\phi_{23} - \frac{5}{16}\phi_{24}$$

$$\phi_3(\xi,\eta) = \phi_4^0 - \frac{1}{16}\phi_{13} + \frac{1}{16}\phi_{14} - \frac{5}{16}\phi_{15} - \frac{5}{16}\phi_{16} + \frac{1}{16}\phi_{17} - \frac{1}{16}\phi_{18}$$

$$\phi_4(\xi,\eta) = \phi_5^0 - \frac{1}{16}\phi_{16} + \frac{1}{16}\phi_{17} - \frac{5}{16}\phi_{18} - \frac{5}{16}\phi_{19} + \frac{1}{16}\phi_{20} - \frac{1}{16}\phi_{21} \qquad \text{(C-3)}$$

$$\phi_5(\xi,\eta) = \phi_3^0 - \frac{15}{16}\phi_{22} - \frac{9}{16}\phi_{23} + \frac{5}{16}\phi_{24},\ \phi_6(\xi,\eta) = \phi_6^0 - \frac{15}{16}\phi_{13} - \frac{9}{16}\phi_{14} + \frac{5}{16}\phi_{15}$$

$$\phi_7(\xi,\eta) = \phi_7^0 - \frac{15}{16}\phi_{16} - \frac{9}{16}\phi_{17} + \frac{5}{16}\phi_{18},\ \phi_8(\xi,\eta) = \phi_8^0 - \frac{15}{16}\phi_{19} - \frac{9}{16}\phi_{20} + \frac{5}{16}\phi_{21}$$

$$\phi_9(\xi,\eta) = \phi_9^0 + \frac{5}{16}\phi_{22} - \frac{9}{16}\phi_{23} - \frac{15}{16}\phi_{24},\ \phi_{10}(\xi,\eta) = \phi_{10}^0 + \frac{5}{16}\phi_{13} - \frac{9}{16}\phi_{14} - \frac{15}{16}\phi_{15}$$

$$\phi_{11}(\xi,\eta) = \phi_{11}^0 + \frac{5}{16}\phi_{16} - \frac{9}{16}\phi_{17} - \frac{15}{16}\phi_{18},\ \phi_{12}(\xi,\eta) = \phi_{12}^0 + \frac{5}{16}\phi_{19} - \frac{9}{16}\phi_{20} - \frac{15}{16}\phi_{21}$$

REFERENCES

1. Tadmor, E. B., Ortiz, M. and Phillips, R. (1996) Philos. Mag. A 73: 1529
2. Miller, R. E. and Tadmor, E. B. (2002) J. Computer-Aided Mater. Design 9: 203
3. Shenoy, V. B., Miller, R., Tadmor, E. B., Rodney D., Phillips, R. and Ortiz, M. (1999) J. Mech. Phys. Solids 47: 611
4. Knap, J. and Ortiz, M. (2001) J. Mech. Phys. Solids 49: 1899
5. Sloan, S.W. (1993) Comput. Struct. 47: 441
6. Park, J. Y. and Im, S. (2008) Phys. Rev. B 77: 184109
7. Lim, J. H., Im, S. and Cho, Y.-S. (2007) Int. J. Numer. Meth. Engng. 72: 835
8. Lim, J. H., Im, S. and Cho, Y.-S. (2007) Comput. Methods Appl. Mech. Engng. 196: 2216
9. Lim, J. H. and Im, S. (2007) Struct. Eng. Mech. 25: 91
10. Pantano, A., Parks, D. M. and Boyce, M. C. (2004) J. Mech. Phys. Solids 52: 789
11. Arroyo, M. and Belytschko, T. (2004) Int. J. Numer. Meth. Engng. 59: 419
12. Qian, D., Wagner, G. J. and Liu, W. K. (2004) Comput. Methods Appl. Mech. Engng. 193: 1603
13. Cousins, C. (1978) J. Phys. C 11: 4867
14. Zanzotto, G. (1996) Acta Crystallogr. A 52: 839
15. Tadmor, E., Smith, G. S., Bernstein, N. and Kaxiras, E. (1999) Phys. Rev. B 59: 235
16. Hughes, T. J. R. (2000) The Finite Element Method. Mineola, New York: Dover Publications, Inc
17. Nocedal, J. (1980) Math. Comput. 24: 773
18. Liu, D. C. and Nocedal, J. (1989) Math. Program. 45: 503
19. Shewchuk, J. R. (1996) "Triangle: 2-D Quality Mesh Generator and Delaunay Triangulator", this code is linked http://www.cs.cmu.edu/∼quake/triangle.html
20. Poncharal, P., Wang, Z. L., Ugarte, D., de Heer, W. A. (1999) Science 283, 1513
21. Tersoff, J. (1988) Phys. Rev. Lett. 61: 2879
22. Brenner, D. W. (1990) Phys. Rev. B 42: 9458
23. Girifalco, L. A., Hodak, M. and Lee, R. S. (2000) Phys. Rev. B 62: 13104
24. Tombler, T. W., Zhou, C., Alexseyev, L., Kong, J., Dai, H., Liu, L., Jayanthi, C. S., Tang, M., Wu, S. (2000) Nature 405: 769
25. Kwon, S., Lee, Y., Park, J. Y., Sohn, D., Lim, J. H., and Im, S. (2009) J. Comput. Phys. 228, 4789
26. Park, J. Y., Park, C.-H., Park, J. S., Kong, K., Chang, H., and Im, S. (2009) J. Mech. Phys. Solids (accepted for publication)

CHAPTER 16

ELECTRONIC PROPERTIES AND REACTIVITIES OF PERFECT, DEFECTED, AND DOPED SINGLE-WALLED CARBON NANOTUBES

WEI QUAN TIAN[1], LEI VINCENT LIU[2], YA KUN CHEN[2], AND YAN ALEXANDER WANG[2]

[1]*State Key Laboratory of Theoretical and Computational Chemistry, Institute of Theoretical chemistry, Jilin University, Changchun 130061, China*

[2]*Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, BC V6T 1Z1, Canada, e-mail: yawang@chem.ubc.ca*

Correspondance to Y.A. Wang, yawang@chem.ubc.ca

Abstract: After we thoroughly surveyed first-principles theoretical methods commonly employed in the studies of carbon nanotubes, we highlighted the performance of such *ab inito* methods on the electronic properties and reactivities of perfect, vacancy-defected, and hetroatom-doped single-walled carbon nanotubes. We have found that a rich chemistry can take place at the vacancy defect and doping sites of nanosystem: this very fact will enable experimental scientists to produce novel functionalized nanosize materials with much higher level of precise control of the manufacturing process

Keywords: Single-walled carbon nanotube, Reactivity, Electronic property, Functionality, Doping, Vacancy defect

16.1. SCOPE

In light of the rapid progress in experimental and theoretical studies of carbon nanotubes, it is very difficult, if not impossible, to write a comprehensive book chapter on the research of carbon nanotubes. Thus, we concentrate on the electronic structures of doped and vacancy-defected single-walled carbon nanotubes (SWCNTs), which have been our major focus in this field [1].

We will first review the properties of SWCNTs and the theoretical methods to study SWCNTs. Then, we report some results from the studies of the electronic structures of perfect SWCNTs, vacancy-defected fullerenes and SWCNTs, and doped SWCNTs, and from modeling of gas adsorptions on doped SWCNTs and chemical reactions involving vacancy-defected SWCNTs.

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 421–471.
DOI 10.1007/978-1-4020-9785-0_16,

16.2. INTRODUCTION

The discovery of SWCNT [2, 3] has spurred extensive experimental and theoretical investigations on the chemical [4–17] and physical [18–30] properties of SWCNTs due to the potential applications [31] of SWCNTs in molecular electronics [32–34], chemical sensor [35–37], vacuum electronic devices [38], field emission flat panel display [39], catalysis [40], optics [41–44], and hydrogen storage materials [45].

According to the wrapping vectors (m,n) [46, 47], a SWCNT, visualized as a roll of graphite sheet, can be classified as metallic and semiconducting. If $n - m = 3q$ (with an integer q), the nanotube is metallic; Otherwise, the nanotube is semiconducting with a narrow or moderate band gap [47]. As rolled one-dimensional graphite sheets, nanotubes display different electronic properties from those of plane graphite sheet and fullerenes: the electronic properties of nanotubes can be controlled by the diameters of the nanotubes due to hybridization effects [23]. The pyramidalization of nanotubes is different from fullerenes and there exists π-orbital misalignment between adjacent pairs of conjugated carbon atoms [4], which renders different reactivity of nanotubes from fullerenes and among nanotubes themselves because of different diameters and chiralty [4, 27, 48]. Chemical reactions can take place on the sidewall of nanotubes [49–61] or at the end of nanotubes [7, 58, 59, 62]. Due to the difficulty in purifying and manipulating SWCNTs, the electronic structure, chemical reactivity, and possible applications of SWCNTs are waiting for further investigation. Especially, the reactivity of SWCNTs is still not clear in spite of numerous attempts and it deserves further studies [63].

Partial destruction of the nanotube structure through vacancy insertion [64–72], doping [73–77], or distortion [78–83] can alter the electronic structure and reactivity of nanotubes [70, 75]. Upon doping, some carbon atoms on a SWCNT are replaced by different atoms, producing a hetero SWCNT (HSWCNT) [73–77]. Most HSWCNTs are doped with main group elements [73–77]; the HSWCNT we studied here is doped with metals Pt, Ni, Pd, Sn, or chalcogens Se and Te. It has been shown that the substitution of a metal atom in fullerenes [78–87] renders the metal as an active center in chemical reactions [88–90]. A good understanding of the electronic structure of SWCNTs and defected SWCNTs offers deep insight into the reactivity of SWCNTs and paves the way for general utilizations of SWCNTs in chemical reactions. As a pseudo-one-dimensional system with a tube structure, a SWCNT with defect can sever as a catalyst for gas- and liquid-phase reactions: the reactants are fed from one end of an open SWCNT, the reaction is catalyzed at the defect site of the SWCNT, and the products are released from the other end of the SWCNT. HSWCNTs can also be used in gas sensors due to the chemical activity of the doped heteroatoms. B- and N-doped HSWCNTs attracted extensive attention due to their potential applications in chemical sensors [75], nanosize electronic and photonic devices with various electronic properties [91]. For the time being, the synthetic methods for these HSWCNTs, such as thermal treatment [92] or chemical vapor deposition [93], can only work under very high temperatures of hundreds and thousands of degrees and the position where the heteroatom is doped cannot be controlled with precision. The feasibility of such synthesis under mild conditions has not been explored theoretically nor reported experimentally [94].

On the other hand, the rapid development of theoretical methods makes computational studies on the structure and property of SWCNTs possible, shedding light on possible applications of SWCNTs. However, this does not claim that theoretical investigations will be easy in any way. The size of SWCNTs lies between small molecule and bulky particle in the range of the so-called nanosize, which is too big for accurate quantum mechanical treatment (for capturing quantum effect) and too small for bulky calculations (for macroscopic properties). The quantum effect of nanosize particles is important in studying their properties; this very fact necessitates the application of quantum mechanical treatment for nanosize materials.

In this book chapter, we will report the structures of HSWCNTs and electronic properties of perfect and vacancy-defected SWCNTs. We will use such knowledge to study gas adsorptions on HSWCNT and reaction of ozone (O_3) with a vacancy-defected SWCNT to gauge the reactivities of these nanosystems, and to simulate chemical reaction of a vacancy-defected SWCNT with nitrogen monoxide (NO) to explore the feasibility of synthesis of an N-doped HSWCNT.

16.3. THEORETICAL METHODS

16.3.1. First-Principles Calculations

Quantum mechanical calculations can be carried out for a system by solving the electronic Schrödinger equation:

$$\mathbf{H_e}\psi = E\psi, \tag{16-1}$$

under the adiabatic approximation, in which the electronic wavefunction is restricted to one electronic surface, and under the Born-Oppenheimer approximation that warrants the separation of nuclear and electronic motions because the nuclei move much slower than the electrons [95–99]. The electronic Hamiltonian $\mathbf{H_e}$,

$$\mathbf{H_e} = \mathbf{T_e} + \mathbf{V_{ne}} + \mathbf{V_{ee}}, \tag{16-2}$$

includes the operators for electronic kinetic energy $\mathbf{T_e}$, nucleus-electron attraction $\mathbf{V_{ne}}$, and electron-electron interaction $\mathbf{V_{ee}}$. The total Hamiltonian $\mathbf{H_{tot}}$ of the system includes $\mathbf{H_e}$ and the operators for nuclear kinetic energy $\mathbf{T_n}$ and nucleus-nucleus interaction $\mathbf{V_{nn}}$,

$$\mathbf{H_{tot}} = \mathbf{H_e} + \mathbf{T_n} + \mathbf{V_{nn}}. \tag{16-3}$$

In most quantum mechanical methods, the nuclei are treated as classical particles and only the electronic wavefunction is solved quantum mechanically. If only fundamental physical constants are used in solving the electronic Schrödinger equation, first-principles calculations emerge. Based on the choice of the basic variational variables, first-principles calculations have two flavors: wavefunction-based *ab initio* methods [95–99] and electron-density-based density-functional theory

(DFT) [100–103]. In wavefunction-based *ab initio* methods, the electronic energy of a system can be expressed as,

$$\mathbf{E_e}[\psi\ ] = \mathbf{T_e}[\psi\ ] + \mathbf{V_{ne}}[\psi\ ] + \mathbf{V_{ee}}[\psi\ ]. \tag{16-4}$$

In the quasi-independent-particle models, e.g. the Hartree-Fock (HF) method [95–99] and the Kohn-Sham (KS) method [100–103], the electronic wavefunction is approximated by a single Slater determinant for an N-electron system,

$$\psi = \frac{1}{\sqrt{\mathrm{N!}}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathrm{N}) & \chi_2(\mathrm{N}) & \cdots & \chi_N(\mathrm{N}) \end{vmatrix}, \tag{16-5}$$

where χ_i is the ith spin orbital. Because of the quasi-independent-particle nature, the instantaneous multi-electron interaction is approximated as if an electron is moving in an average electron–electron potential due to other electrons in the system. The variational HF (or KS) equation is solved self-consistently with a basis set representation for molecular orbitals (MOs), which are usually constructed from the linear combination of atomic orbitals (LCAOs) [95−104]. The HF approximation is not capable of describing highly correlated systems, in which the instantaneous electron–electron interaction is crucial to a correct description of the electronic structure. High-level methods, such as multiconfiguration self-consistent field (MCSCF) [105], multireference configuration interaction (MRCI) [106], or coupled-cluster (CC) methods [107], are then necessary for treating electron correlation more accurately. However, the prohibitive computing resource requirements of such high-level methods prevent their general applications in large systems [95–99]. Even for the HF approximation, the system size that can be treated is still moderate, since the computational cost formally scales as the fourth power of the number of basis functions. Other high-level methods, such as CC, scale even worse with respect to the system size. One solution to reduce the computing cost is developing linear-scaling methods. Another way is reducing the computational cost by approximating the most time-consuming step of the calculation, the evaluation of two-electron integrals. This latter approximation results in semiempirical methods [95–99].

16.3.2. Semiempirical Quantum Mechanical Methods

In the HF approximation, the electron-electron interaction is evaluated through the operator $\mathbf{V_{ee}}$,

$$\mathbf{V_{ee}} = \sum_{\mathrm{i}}^{\mathrm{N}} \sum_{\mathrm{j>i}}^{\mathrm{N}} \frac{1}{|\mathrm{r_i} - \mathrm{r_j}|}, \tag{16-6}$$

which results in two-electron integrals. Through linear combination of atomic basis functions $\{\varphi_a\}$, molecular orbitals $\{\chi_i\}$ can be written as,

$$\chi_i = \sum_a c_{ai}\varphi_a, \tag{16-7}$$

where $\{c_{ai}\}$ are the MO coefficients. Following the variation principle, the HF equation can be derived directly from Eqs. (16-1), (16-2), and (16-5) as a pseudo one-electron eigen-equation:

$$\mathbf{F}\chi_i = \varepsilon_i\chi_i, \tag{16-8}$$

where **F** is the Fock operator. In the atomic basis function representation, the HF equation is recasted as the Roothaan-Hall equation,

$$\mathbf{FC} = \mathbf{SC}\varepsilon, \tag{16-9}$$

where **F** is the Fock operator matrix, **C** is the MO coefficient matrix, and **S** is the overlap matrix, respectively. The Fock operator **F** could be written as a sum of one-electron operator **h** and two-electron operators,

$$\mathbf{F} = \mathbf{h} + \sum_j^N (J_j - K_j), \tag{16-10}$$

where J_j is the Coulomb operator and K_j is the exchange operator for two-electron interactions, respectively. Most of the CPU time is spent in calculating the two-electron integrals due to the two-electron operators J_j and K_j [95–99].

In semiempirical methods [97−99], only valence electrons are taken into account explicitly; the core electrons are implicitly included in the nuclear part by reducing the nuclear charge. The evaluation of the two-electron integrals for valence electrons is simplified by the introduction of parameters. To further reduce the computational cost, only a minimum number of basis functions are introduced to describe the motion of electrons in semiempirical methods. The basic approximation in semiempirical methods is the zero differential overlap (ZDO) approximation, in which all the products of basis functions of the same electron on different atoms are neglected. The remaining integrals are parameterized and fitted through benchmarking with available experimental data. Different semiempirical methods can be designed depending on how the approximations in the neglect of two-electron integrals and in the fitting of parameters are made [97−99]. Among all variants of the ZDO approximation, the complete neglect of differential overlap (CNDO) method [108–110] is the crudest approximation. In the CNDO method, only Coulomb one-center and two-center two-electron integrals remain. With more refined ZDO approximation, the intermediate neglect of differential overlap (INDO) method [111] and the neglect of diatomic differential overlap (NDDO) methods [108–110] (including MNDO [112], AM1 [113], and PM3 [114, 115]) came with much

improved accuracy. In present work, the PM3 method is employed in combined quantum chemical calculations.

16.3.3. Density-Functional Theory

In conventional *ab initio* methods, the computational cost scales formally at least as the fourth power of the number of basis functions. To incorporate electron correlation effects into the HF approximation, high-level methods have to pay a big price in terms of computational cost. On the other hand, the electron-density-based DFT has a favorable scaling factor [95–103]: the third power of the number of basis functions in the KS scheme, and more importantly, the electron correlation effects are taken into account through correlation functionals in DFT [100–103].

In DFT, the electronic energy is expressed as a sum of various energy density functionals,

$$E_e[\rho] = T_s[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho], \tag{16-11}$$

where ρ is the single electron density, $T_s[\rho]$ is the electronic kinetic energy, $E_{ne}[\rho]$ is the nuclear-electron attraction energy, $J[\rho]$ is the electron–electron Coulomb interaction energy, and $E_{xc}[\rho]$ is the exchange and correlation energy. $E_{xc}[\rho]$ can be further split into two pieces: the exchange energy $E_x[\rho]$ and the correlation energy $E_c[\rho]$.

$E_{xc}[\rho]$ is mainly modeled in three ways: the local-density approximation (LDA), the generalized gradient approximation (GGA), and the hybrid approach [100–103, 116]. In the LDA, the energy of a system depends on the local value of the electron density [100–103]. In the GGA, the energy of a system depends not only on the local value of the electron density but also on the gradients of the electron density of different orders [103]. While in the hybrid DFT, the exact exchange defined in terms of the KS orbitals (just like the one used in the HF approximation defined in terms of the HF orbitals) is admixed with the approximate exchange functional [117]. Such a hybrid approach of mixing the exact exchange further improves the predicative accuracy of DFT method in physical and chemical applications [117].

However, the system size and scaling factor of computational cost limit the application of quantum mechanical methods in the studies of nanoparticles. To overcome this difficulty, the ONIOM model [118] has been developed to be a compromise of the computational feasibility and accuracy.

16.3.4. ONIOM Model

In most cases, chemical reactions are localized in the proximity of the active site. Such a localized nature of chemical reactions legitimizes an accurate treatment of the active site with high-level method while the surroundings (the rest of the chemical system) can be treated with low-level theory [118–123]. This embedding approach requires relative low computing resources and still keeps the essential environmental effects exerted on the active site by the surroundings [119–123].

The ONIOM model [118] is a popular one among many methods [119–123] suitable for such applications. The basic formula for ONIOM can be written as [118],

$$E_{total} = E_{RL} - E_{ML} + E_{MH}, \tag{16-12}$$

where E_{total} is the total energy of the system, E_{RL} is the energy of the real system at low-level theory, E_{ML} is the energy of the model system (the active site) at low-level theory, and E_{MH} is the energy of the model system at high-level theory. The gradient and the second-order derivatives of the total energy with respect to the nuclear coordinates can be calculated in a similar manner with the help of link atoms [118]. A link atom is a buffer atom in the model system replacing the actual atom in the real system connected to the model system.

For very large systems, the low-level theory for the overall real system is usually molecular mechanics. In molecular mechanics [95–99], the electronic structure of the system is not explicitly considered. The interatomic interactions are divided into bond, angle, torsion angle, and weak interactions (including the van der Waals interactions). Such interactions are parameterized through fitting to the available experimental data or any high-level quantum mechanical predictions. Depending on the fitting procedure, different parameterizations give rise to various force fields for the interatomic interactions. Amber [119–124], Charmm [125], MM3 [126], and UFF [127] are the most popular force fields among many [100–103].

16.3.5. Molecular Dynamical Simulations

A normal quantum mechanical study on the potential energy surface (PES) investigates the reaction pathway for a particular reaction channel involving stationary points on the PES. What we can get from such a study is the structures, properties, and energetics of the stationary points from which we can calculate reaction energies, barriers, and rates. On the other hand, molecular dynamics (MD) involving propagation of nuclei in molecules on the PES by solving Newton's equation of motion provides rich information about reactivity and dynamics of the system, so that the motions of the component atoms during the propagation and the instant interactions among atoms can be revealed. Though the PES can be obtained by fitting to experimental or computational data, the fitting is not a trivial task even if the necessary data are available. Theoretical MD simulation, an alternative means to study molecular propagation and chemical reactions, provides complementary information about the thermodynamical and dynamical properties and microscopic motions of nuclei in a chemical reaction.

The force field, which confines the nuclear motions, is important as it dicates the quality of the MD simulation. Despite that empirical force fields have gained wide popularity in MD simulations of large systems in biology [128, 129], solid state physics, and surface science [130], high-quality force field from quantum mechanics is necessary for accurate description of quantum effects. Therefore,

quantum mechanical MD (QMMD), in which the force field is computed quantum mechanically, is a natural choice for this purpose [131−136].

For large systems, a dilemma arises: modeling quantum effects is important but accurate quantum mechanical methods cannot handle such large systems with present computational facility. To overcome this obstacle, we utilize the ONIOM model [118], in which the most important forces are computed by quantum mechanical method and the remaining system is described by molecular mechanical force fields.

16.4. SINGLE-WALLED CARBON NANOTUBES

SWCNT could be either end-capped or with open end during fabrication [38, 137]. With spherical curvature, the capping ends can have different electronic structure from that of the sidewall of the SWCNT and modify the electronic properties of the SWCNT around the ends [18, 68]. We studied two models for SWCNT: one with hemispherical capping ends (a SWCNT rod) and the other with open ends saturated by hydrogen atoms (a SWCNT clip). The model for SWCNT is (5,5) SWCNT.

16.4.1. Perfect SWCNT Rods

The caps in the hemispherically capped nanorod are half spheres of a fullerene, which vary as the diameter and the structure of the SWCNT change [17, 22, 64]. The caps in capped nanorod were predicted to play an important role in the electronic property of the nanorod [11]. The caps host some localized electronic states [18, 138], because of the relatively unstable pentagons presented there [139]. However, the caps do not contribute significantly to the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) of the nanorods [10]. Differences in curvature and π bonding distinguish the chemical properties of fullerenes and SWCNTs [4] and thus divide the nanorod into at least two regions: the caps and the sidewall.

Using Gaussian 03 quantum chemical program package [140] and Pople's 6-31G Gaussian basis set [141–144], we studied two nanorods: C_{170} with D_{5h} symmetry and C_{180} with D_{5d} symmetry. The geometries of these nanorods were optimized by GGA-type DFT based method (BPW91) with Becke's exchange functional (B) [145] and Perdew's correlation functional (PW91) [146, 147]. Natural bond orbitals (NBOs) [148] were calculated to estimate partial charges of atoms. The structures, density of states (DOS), and local density of states (LDOS) of the two nanorods are shown in Figure16-1.

The overall DOS and LDOS of these two SWCNT rods are very similar. The most noticeable feature of the DOS of C_{170} is the separation of peaks at 5.2 eV whereas it is continuous for C_{180}. The similarity of the DOS and LDOS of C_{170} and C_{180} is expected, since only one additional circular *cis*-polyene chain does not change the electronic structure significantly from C_{170} to C_{180}. Previous studies also found the similarity in the HOMO-LUMO gaps of the (5,5) SWCNT rods,

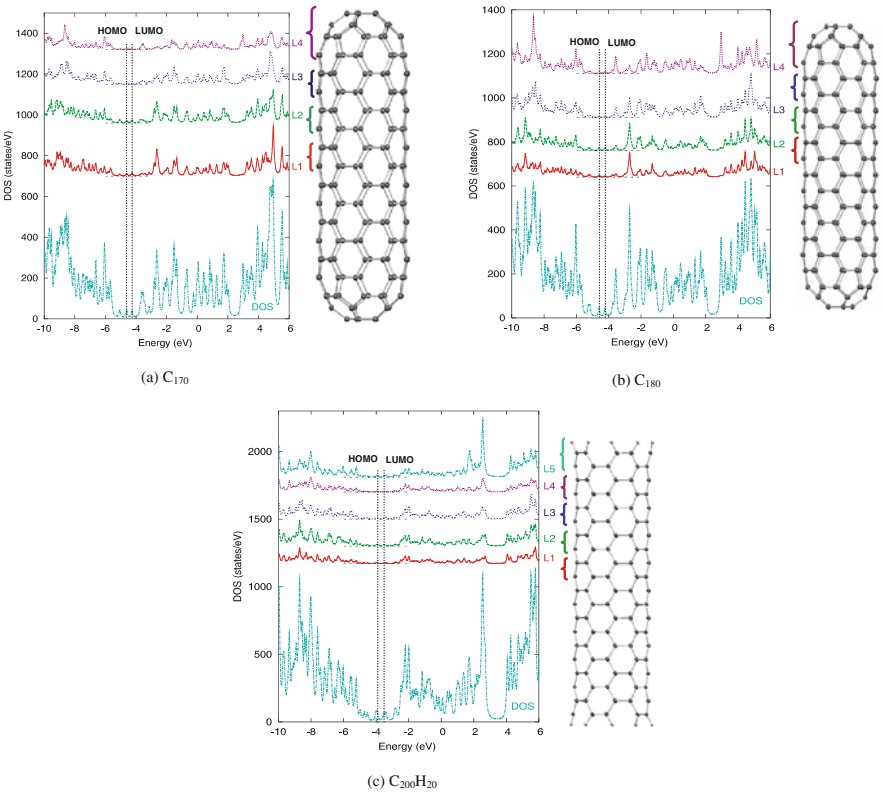


Figure 16-1. (**a**) The density of states and local density of states for the nanorod C_{170} with D_{5h} symmetry. The HOMO has orbital energy −4.61 eV, and the LUMO has orbital energy −4.26 eV. (**b**) The density of states and local density of states for the nanorod C_{180} with D_{5d} symmetry. The HOMO has orbital energy −4.60 eV, and the LUMO has orbital energy −4.23 eV. (**c**) The density of states and local density of states for the open-end (5,5) SWCNT segment $C_{200}H_{20}$ with D_5 symmetry. The HOMO has orbital energy −3.93 eV, and the LUMO has orbital energy −3.52 eV. L1−L5 are the local density of states for each specified layer as outlined on the structures. This figure is adopted from Refs. [94, 180]

C_{170} and C_{180} [11]. The HOMO-LUMO gaps of the (5,5) SWCNT rods C_{170} and C_{180} are smaller than that of the longer open-end (5,5) SWCNT segment (this will be elaborated in the next section). The LDOS shows the contribution of a particular group of atoms to the overall DOS. Figure 16-1a and b display the LDOS of *cis*-polyene chains in C_{170} and C_{180} along the SWCNT axis and the LDOS of the cap (a hemisphere of C_{60}). The shapes of the LDOS of different layers are similar at the frontier molecular orbital (FMO) region, which indicates the delocalization of the FMOs of the SWCNT rod. The contributions to the HOMO, the LUMO, and other occupied FMOs from the caps are not significant; major contributions to the HOMO and the LUMO of C_{170} and C_{180} are from the sidewall of the SWCNTs. Figure 16-1a and b clearly indicate that the conspicuous contributions to the DOS

of the SWCNT rod from the LDOS of the caps lie about 1.0 eV below the HOMO and 0.5 eV above the LUMO.

MOs give detailed information about the contributions of the LDOS from each layer to the DOS of the SWCNT rod. Some occupied and unoccupied FMOs for SWCNT rods C_{170} and C_{180} are plotted in Figure 16-2. The highest four occupied MOs of C_{170} and C_{180} are delocalized π orbitals with contributions from the sidewall of the SWCNT rod. The occupied MOs with major contributions from the caps lie about 1.0 eV below the HOMO, as also manifested by the LDOS of the cap in Figure 16-1. The HOMO and the LUMO have sole contributions from the sidewall of the SWCNT rod. The lowest two unoccupied MOs of C_{170} and C_{180} are also π orbitals delocalized on the sidewall of the SWCNT rod, and the next four unoccupied MOs are two two-fold degenerated localized MOs on the caps. The patterns of the HOMO and the LUMO in C_{170} are different from their counterparts in C_{180}. Such pattern change was also observed for shorter SWCNT rods before [10].

From the MOs, one can infer that, when reacting with strong electron acceptors, the SWCNT rods C_{170} and C_{180} will donate electrons from the sidewall to the

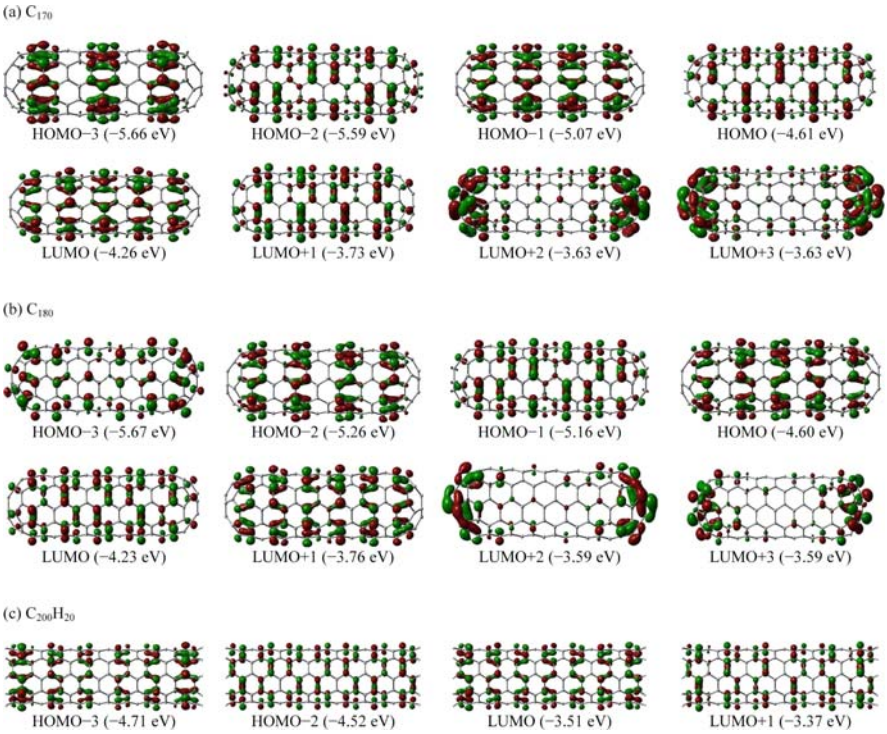


Figure 16-2. The frontier molecular orbitals of (**a**) the nanorod C_{170} with D_{5h} symmetry, (**b**) the nanorod C_{180} with D_{5d} symmetry, and (**c**) the open-end (5,5) SWCNT segment $C_{200}H_{20}$. HOMO−*n* (*p* eV) is the *n*th orbital below the HOMO with orbital energy *p* eV. LUMO+*m* (*q* eV) is the *m*th orbital above the LUMO with orbital energy *q* eV. This figure is adopted from Refs. [94, 180]

electron acceptors. According to the NBO analysis, when C_{170} and C_{180} accept electrons, the first four electrons will go to the middle of the sidewall of the SWCNT rod, and any extra (up to eight) electrons will go to the caps. From the MOs of C_{170} and C_{180}, one cannot see clearly the separation of the cap from the sidewall, though there are some gradual geometric changes from the cap to the sidewall [149]. The delocalized MOs on the sidewall extend to the ridge of the pentagons of the caps. The pentagon regions have 6–6 (between two hexagons) and 6–5 (between a hexagon and a pentagon) CC bond alternation similar to that in C_{60}. The atoms on the last layer of L4 (in Figure 16-1a and b), bridge between the sidewall and the cap of the SWCNT rod, have the largest negative charges in C_{180} and large negative charges in C_{170}. In chemical reactions, this layer is reactive toward electron acceptors. The middle layers of the sidewall have positive charges.

16.4.2. Open-End SWCNT Segment

Open-end SWCNTs and end-capped SWCNTs with vacancy defects of the similar length can have very different electronic structures and reactivity. To understand such differences, we studied an open-end (5,5) SWCNT segment $C_{200}H_{20}$, which has 200 carbon atoms, saturated with 20 hydrogen atoms at the two open ends (Figure 16-1c). The same method and basis set used for C_{170} and C_{180} were employed to study $C_{200}H_{20}$. The HOMO-LUMO gap of the $C_{200}H_{20}$ segment (0.41 eV) is larger than those of the SWCNT rods, C_{170} (0.35 eV) and C_{180} (0.37 eV). Both of the HOMO and the LUMO are destabilized when compared with those of C_{170} and C_{180}, in spite that the $C_{200}H_{20}$ segment is longer than the SWCNT rods C_{170} and C_{180}. The stabilization effect of the hemispherical caps to the SWCNT rod is evident from the comparison of MO energies of the open-end SWCNT segments and SWCNT rods. The bands in the DOS of $C_{200}H_{20}$ (Figure 16-1c) shift to higher energy regions than those of the SWCNT rod C_{180} (Figure 16-1b). The LDOS of the circular *cis*-polyene chains of $C_{200}H_{20}$ are very similar. Figure 16-2c shows that each layer contributes roughly equally to the FMOs and there is no localized state on $C_{200}H_{20}$. The HOMO and the LUMO of short (5,5) SWCNT segments with open ends were studied before and showed alternating nodal pattern [10], which is not observed here for longer (5,5) SWCNT segments.

Overall, the hemispherical fullerene caps introduce localized states and stabilize the SWCNT rod.

16.5. VACANCY-DEFECTED FULLERENES AND SWCNTS

Removal of one carbon atom from different positions of the capping ends of a SWCNT rod results in different vacancy-defected SWCNTs. For simplicity, we studied single-vacancy-defected fullerenes C_{59} and C_{69} (originated from C_{60} and C_{70}, respectively), as models for the end caps of SWCNT rods. For vacancy-defected SWCNT with defects at the middle of the sidewall, (5,5) and (10,0) SWCNTs were employed as models.

16.5.1. Vacancy-Defected Fullerenes

Defected C_{60} (and possibly C_{70}) with odd number of vacancies have been produced through laser desorption ionization of $C_{60}O$ [150]. Vacancies on defected fullerenes were proposed to serve as the windows for atoms or small molecules to enter the cage in the endohedral fullerene chemistry [151]. Vacancy-defected fullerene C_{59} has been investigated theoretically [152−159] but with some contradictory results [156−159]. It was reported [156, 157] that triplet C_{59}(5-8) [with a pentagon adjacent to an octagon] is more stable than C_{59}(4-9) [with a tetragon adjacent to a nonagon] (Figure 16-3b and c). The initial singlet C_{59}(4-9) isomer transforms into the singlet C_{59}(5-8) structure during a geometry optimization based on spin-unrestricted DFT method [157]. However, Ribas-Ariño and Novoa [158] found that both C_{59}(5-8) and C_{59}(4-9) have singlet ground states. Using tight-binding molecular-dynamics and *ab initio* methods, Andriotis et al. studied single-vacancy

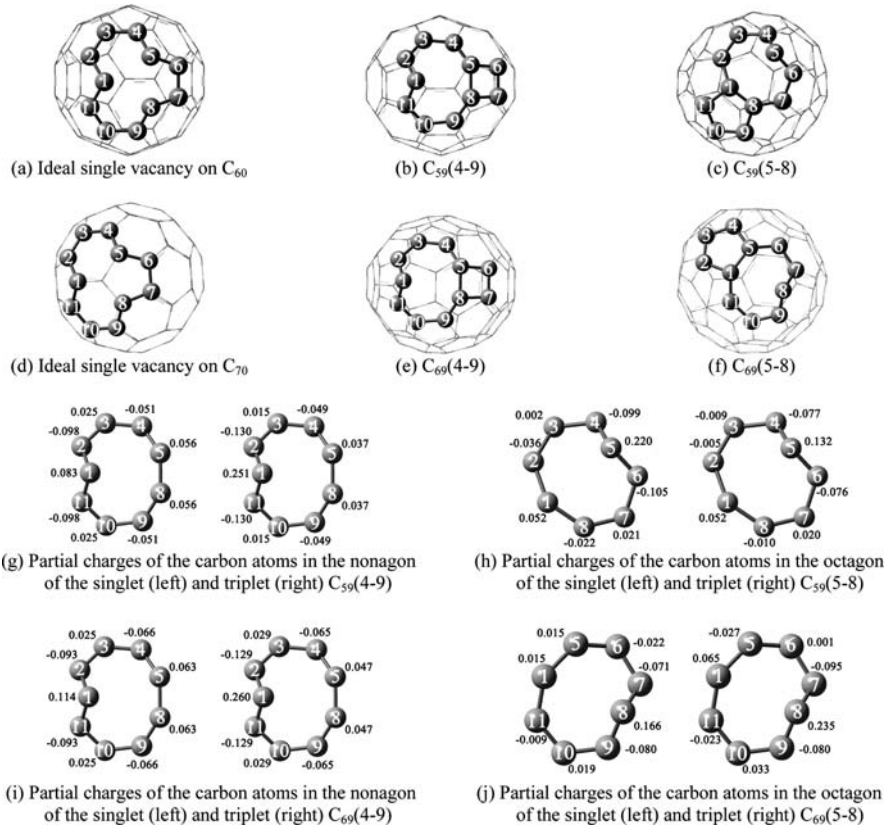


Figure 16-3. Structures of the single-vacancy defected C_{60} and C_{70} with partial charges of the important carbon atoms

defected C_{60} in a C_{60} polymer and found that two out of the three dangling bonds of the ideal single-vacancy defect do not recombine [159]. Thus, we reexamined the electronic properties of C_{59}(5-8) and C_{59}(4-9) to resolve these controversies and to explore the isomerization pathways between these two isomers. To compare with the C_{60} system, we also study the vacancy-defected C_{70}. Structurely, C_{70} has one more hexagonal belt between two hemispheres of C_{60}, and vacancy-defected C_{70} might have some similar properties to vacancy-defected C_{60} but with different structures.

The hybrid DFT method, B3LYP [117, 160], with the standard 6-31G(d) basis set was used for geometry optimizations. Single-point calculations at the B3LYP/6-311G(d) level of theory were performed with the B3LYP/6-31G(d) optimized geometries. NBO analysis [148] was performed to determine the partial charge distribution and the bonding characters of these systems. B3LYP/6-31G was employed to study the isomerization between C_{59}(5-8) and C_{59}(4-9) on the singlet and triplet PESs. Spin-unrestricted calculations were carried out for the singlet and triplet electronic states, and spin-restricted calculations were also done for the singlet electronic states. In light of recent numerical tests [161], the convergence criteria of all calculations were set very tight (e.g., 10^{-8} for the density convergence) to ensure full convergence.

The two types of bonds in C_{60}, the shorter double bonds and the longer single bonds, are predicted to be 1.395 and 1.453 Å at the B3LYP/6-31G(d) level of theory, respectively, in good agreement with the experimental data (1.390 and 1.453 Å, respectively) [162]. The structures of C_{60} and C_{70} with an ideal single vacancy and related point defects are shown in Figure 16-3, in which important carbon atoms are labelled numerically. In C_{59}, C1 forms a bond with C5 or C8 leading to the C_{59}(5-8) isomer (5 and 8 denoting the newly forming pentagon and octagon), or C5 forms a bond with C8 leading to the C_{59}(4-9) isomer (4 and 9 denoting the newly forming tetragon and nonagon). In C_{69}, C1 forms a bond with C5 or C8 leading to the C_{69}(5-8) isomer, or C5 forms a bond with C8 leading to the C_{69}(4-9) isomer. In each isomer, there is one unsaturated carbon atom in a pentagon for the (4-9)-type defects or in a hexagon for the (5-8)-type defects.

In comparison with the HF/3-21G and B3LYP/3-21G results for the C_{59}(4-9) isomer [156], the B3LYP/6-31G(d) calculations predict shorter bond lengths for most bonds due to the larger basis set used. Table 16-1 lists the total and cohesive energies of the isomers computed at the B3LYP/6-311G(d)//B3LYP/6-31G(d) level of theory. According to the cohesive energies, the triplet C_{59}(5-8) and C_{69}(5-8) isomers are the most stable isomers for C_{59} and C_{69}, respectively. However, the small spin contamination in the spin-unrestricted calculations may artificially stabilize the triplet state over the closed-shell singlet state. With such small energy difference, the singlet and triplet electronic states of C_{59} and C_{69} have similar stabilities within present DFT treatment.

Reference [158] reported a higher spin-contamination in the spin-unrestricted calculation of singlet C_{59}(5-8) and the transformation of the open-shell singlet C_{59}(4-9) structure into the open-shell singlet C_{59}(5-8) structure. We will focus on the singlet and triplet potential energy curves for the isomerization of these two

Table 16-1. Energies of C_{60}, C_{59}(4-9), C_{59}(5-8), C_{70}, C_{69}(4-9), and C_{69}(5-8) at the B3LYP/6-311G(d)//B3LYP/6-31G(d) level of theory

	E_{total} (Hartree)		E_{sa}[c] (kcal/mol)		ΔE[d] (kcal/mol)	
Model	S (R/U)[a]	T (U)[a]	S (R/U)	T (U)	S (R/U)	T (U)
C_{60}	−2286.5888	–	159.27	–	–	–
C_{59}(4-9)	−2248.2431	−2248.2418 (2.05)[b]	156.76	156.75	2.51	2.52
C_{59}(5-8)	−2248.2690	−2248.2744 (2.05)	157.04	157.09	2.23	2.18
C_{70}	−2667.7857	–	160.16	–	–	–
C_{69}(4-9)	−2629.4433	−2629.4396 (2.04)	158.05	158.02	2.11	2.14
C_{69}(5-8)	−2629.4629	−2629.4670 (2.05)	158.23	158.27	1.93	1.89

[a] R: Spin-restricted; U: Spin-unrestricted; S: Singlet; T: Triplet.
[b] The values in the parentheses are the $\langle S^2 \rangle$ values.
[c] The stability energy per atom, defined as E_{ts}/n, where E_{ts} is the energy difference between the isolated carbon atoms and the cluster and n is the total number of carbon atoms in the cluster.
[d] $\Delta E = E_{sa}$ (perfect fullerene) − E_{sa} (defected fullerene).

structures. For singlet C_{59}(4-9), the dangling carbon atom C1 forms single bonds with C2 and C11. The bond lengths of C1−C2 and C1−C11 are both 1.441 Å because of C_s symmetry. Based on the NBO analysis, there are two 2-center σ NBOs involving C1 and two lone-pair-type NBOs for C1. The partial charges of the carbon atoms in the nonagon of singlet and triplet C_{59}(4-9) are shown in Figure 16-3g. C1 has +0.08 charge and its two neighbours have −0.10 charges in the singlet case, whereas in the triplet case, C1 has a much larger positive charge (+0.25) and its two neighbours also have larger negative charges (−0.13). The relevant FMOs of singlet C_{59}(4-9) are shown in Figure 16-4A. Quite interestingly, there is nearly no electron population on C1 in the HOMO. The first orbital below the HOMO (HOMO−1)[1] is mainly the lone-pair $sp^{1.50}$ hybridized orbital. The LUMO is mainly the lone-pair *p* unhybridized orbital. The HOMO energy of singlet C_{59}(4-9) is −6.15 eV, higher than that of perfect C_{60} (−6.40 eV). The LUMO energy of singlet C_{59}(4-9) is −4.22 eV, lower than that of perfect C_{60} (−3.68 eV). Thus, the HOMO-LUMO gap is only 1.93 eV, smaller than that of perfect C_{60} (2.72 eV).

For triplet C_{59}(5-8), the dangling carbon atom C5 forms a single bond with C4 and a double bond with C6. The bond lengths of C5−C4 and C5=C6 are 1.399 and 1.357 Å, respectively. The point group of triplet C_{59}(5-8) is C_1. Based on the NBO analyses, there are three 2-center NBOs involving C5 and two lone-pair-type NBOs for C5. The alpha-spin FMOs are shown in Figure 16-4B. The HOMO contains the lone-pair *p* unhybridized orbitals (Figure 16-4Bc). The orbital energy of the alpha-spin HOMO of triplet C_{59}(5-8) is −5.78 eV, higher than that of perfect C_{60} (−6.40 eV). The orbital energy of the alpha-spin LUMO of triplet C_{59}(5-8)

[1] Throughout the text, HOMO−*n* and LUMO+*m* denote the *n*th molecular orbital below the HOMO and the *m*th molecular orbital above the LUMO, respectively.

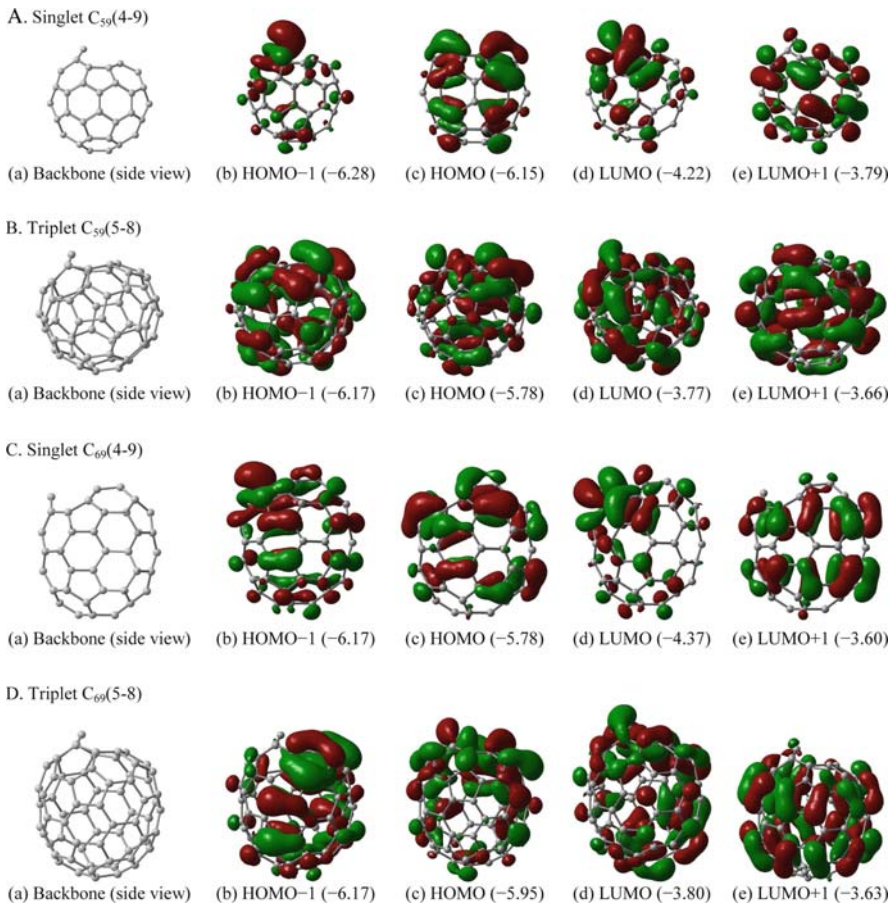


Figure 16-4. (**A**) Frontier molecular orbitals of singlet C_{59}(4-9); (**B**) alpha-spin frontier molecular orbitals of triplet C_{59}(5-8); (**C**) frontier molecular orbitals of singlet C_{69}(4-9); (**D**) alpha-spin frontier molecular orbitals of triplet C_{69}(5-8). HOMO−*n* (*p*) is the *n*th molecular orbital below the HOMO with orbital energy *p* eV. LUMO+*m* (*q*) is the *m*th molecular orbital above the LUMO with orbital energy *q* eV

is −3.77 eV, slightly lower than that of perfect C_{60} (−3.68 eV). The orbital energy of the beta-spin HOMO of triplet C_{59}(5-8) is −6.10 eV, higher than that of perfect C_{60} (−6.40 eV). The orbital energy of the beta-spin LUMO of triplet C_{59}(5-8) is −4.41 eV, lower than that of perfect C_{60} (−3.68 eV). Thus, the alpha-spin HOMO-LUMO gap (2.01 eV) and the beta-spin HOMO-LUMO gap (1.69 eV) are both smaller than that of perfect C_{60} (2.72 eV).

For singlet C_{69}(4-9), the dangling carbon atom C1 forms single bonds with C2 and C11. The bond lengths of C1−C2 and C1−C11 are both 1.434 Å because of C_s symmetry. The partial charges of the carbon atoms in the nonagon are shown in Figure 16-3i for singlet and triplet C_{69}(4-9). C1 has +0.14 charge and its two neighbours have −0.09 charges in the singlet case, whereas in the triplet case,

C1 has +0.26 charge and its two neighbours have −0.13 charges. The relevant FMOs in singlet C_{69}(4-9) are shown in Figure 16-4C. The LUMO (Figure 16-4Cd) contains the lone-pair unhybridized *p* orbital. The HOMO energy of singlet C_{69}(4-9) is –6.22 eV, slightly higher than that of perfect C_{70} (−6.34 eV). The LUMO energy of singlet C_{69}(4-9) is −4.37 eV, lower than that of perfect C_{70} (−3.67 eV). Thus, the HOMO-LUMO gap is only 1.85 eV, smaller than that of perfect C_{70} (2.67 eV).

For triplet C_{69}(5-8), the dangling carbon atom C8 forms a single bond with C7 and a double bond with C9. The bond lengths of C8−C7 and C8=C9 are 1.397 and 1.358 Å, respectively. The HOMO contains the lone-pair *p* unhybridized orbital of C1 (Figure 16-4Dc). The orbital energies of the alpha- and beta-spin HOMO of triplet C_{69}(5-8) are higher than that of C_{70} while the orbital energies of LUMO are lower than C_{70}. Thus, the alpha-spin HOMO-LUMO gap (2.15 eV) and the beta-spin HOMO-LUMO gap (1.44 eV) are both smaller than that of perfect C_{70} (2.67 eV).

Andriotis et al. studied the single-vacancy defected C_{60} in a C_{60} polymer by using tight-binding molecular-dynamics and *ab initio* methods [159]. In their study, two out of the three dangling bonds of the ideal single-vacancy defect do not recombine, about +0.5 charge resides on the three carbon atoms with dangling bonds, and the spin polarization occurs at the σ dangling bonds. In our study of the isolated single-vacancy defected C_{60}, two out of the three dangling bonds do recombine. Only about +0.2 charge accumulates at the carbon atom with dangling bond and the two recombined carbon atoms. The spin polarization mainly localizes at the remaining carbon atom with dangling bond for both triplet C_{59}(5-8) and triplet C_{59}(4-9). The different conclusions might be caused by the different environments modeled by Andriotis et al. (in a C_{60} polymer) [159] and by us (in gas phase). More refined theoretical treatment should be carried out to confirm the findings of Andriotis et al. [159].

It is interesting to recognize that these vacancy-defected fullerenes can be actually treated as carbenes. A carbene is a divalent carbon atom with four valence electrons, and its two nonbonding electrons can lead either to singlet state or to triplet state. The simplest carbene is methylene. If methylene is linear, it will have two degenerate *p* orbitals, and each of the two nonbonding electrons will occupy one of these two *p* orbitals with the same spin, thus yielding a triplet ground state. If methylene is bent, the degeneracy of these two *p* orbitals is destroyed. The orbital perpendicular to the bent methylene is called "*p*", and the other orbital is called "σ" that hybrids with the *s* orbital and becomes stabilized. The more *s* character this σ orbital has, the bigger the energy gap between the σ and *p* orbitals will be. If the σ–*p* energy gap is big, the two nonbonding electrons will prefer to stay in the *p* orbital with opposite spins, thus becoming a singlet carbene. If the σ–*p* energy gap is small, the two nonbonding electrons will prefer to stay in different orbitals with the same spin, thus producing a triplet carbene.

The results indicate that the singlet carbene prefers to stay in the pentagon and the triplet carbene prefers to stay in the hexagon of the isomers of C_{59} and C_{69}. One reason for this scenario is that the triplet carbene prefers bigger bond angle in

the hexagon of the defect site, whereas the singlet carbene prefers the smaller bond angle in the pentagon of the defect site. Another reason is due to different electronic effects in C_{59}(4-9) and C_{59}(5-8) [163]. From the NBO analyses, we know that the σ orbitals of the carbenes of C_{59}(4-9) and C_{59}(5-8) are $sp^{1.50}$ and $sp^{3.57}$ hybridized, respectively. This means that the σ–*p* energy gap of the carbene of C_{59}(4-9) is larger than that of C_{59}(5-8). On the other hand, we find that the *p* orbital of the carbene of C_{59}(5-8) forms a π bond with one of its neighbour carbon atoms, whereas this is not true for the carbene of C_{59}(4-9). This effect can lower the energy of the *p* orbital of the carbene of C_{59}(5-8), thus leading to a smaller σ–*p* energy gap, which also explains the larger stability of triplet C_{59}(5-8) and triplet C_{69}(5-8).

The isomerization pathways of C_{59}(4-9) and C_{59}(5-8) isomers on the singlet and triplet PESs are further explored. The relative energies, the transition-state structures, and the imaginary vibrational modes are shown in Figure 16-5. The energy of triplet C_{59}(5-8) is set as the reference zero-point. The barrier of singlet C_{59}(4-9) transforming into singlet C_{59}(5-8) is 35.69 kcal/mol and the reverse barrier is 54.50 kcal/mol. The barrier between singlet C_{59}(5-8)L (L denoting C1−C3 bond-formation) and singlet C_{59}(5-8)R (R denoting C2−C3 bond-formation) is 49.45 kcal/mol. The barrier of triplet C_{59}(4-9) transforming into triplet C_{59}(5-8) is 17.49 kcal/mol and the reverse barrier is 41.69 kcal/mol. The barrier between triplet C_{59}(5-8)L and triplet C_{59}(5-8)R is 38.87 kcal/mol. Clearly, the isomerization takes place more readily on the triplet PES. Under the influence of strong external field or irradiation impact during the vacancy defect creation [150, 164–167], such isomerization can take place feasibly. Both the relative energies of C_{59}(4-9) and C_{59}(5-8) and the isomerisation barrier between C_{59}(4-9) and C_{59}(5-8) make C_{59}(5-8) the major component during C_{59} formation.

Vertical and adiabatic values of the electron ionization and affinity energies (without the zero-point correction) for ground-state C_{59}(4-9), C_{59}(5-8), C_{69}(4-9), and C_{69}(5-8) are shown in Table 16-2. The vertical electron affinity (VEA) or vertical detachment affinity is defined as the energy difference between the neutral cluster and its anion both at the equilibrium geometry of the anion [168]. The adiabatic electron affinity (AEA) or simple electron affinity is defined as the energy difference between the neutral cluster and its anion at their own equilibrium geometries [168]. The vertical ionization potential (VIP) is defined as the energy difference between the cation and its neutral cluster both at the equilibrium geometry of the neutral cluster. The adiabatic ionization potential (AIP) is defined as the energy difference between the cation and its neutral cluster at their own equilibrium geometries.

The VEA and AEA of the C_{59}(4-9) and C_{69}(4-9) clusters are smaller than those of the C_{59}(5-8) and C_{69}(5-8) clusters, whereas the VIP and AIP of the C_{59}(4-9) and C_{69}(4-9) clusters are larger than those of the C_{59}(5-8) and C_{69}(5-8) clusters. The differences between the VEA and the AEA and between the VIP and the AIP for all these clusters are very small, indicating that the optimized geometries of the neutral clusters and their corresponding anionic and cationic clusters are close to one another.

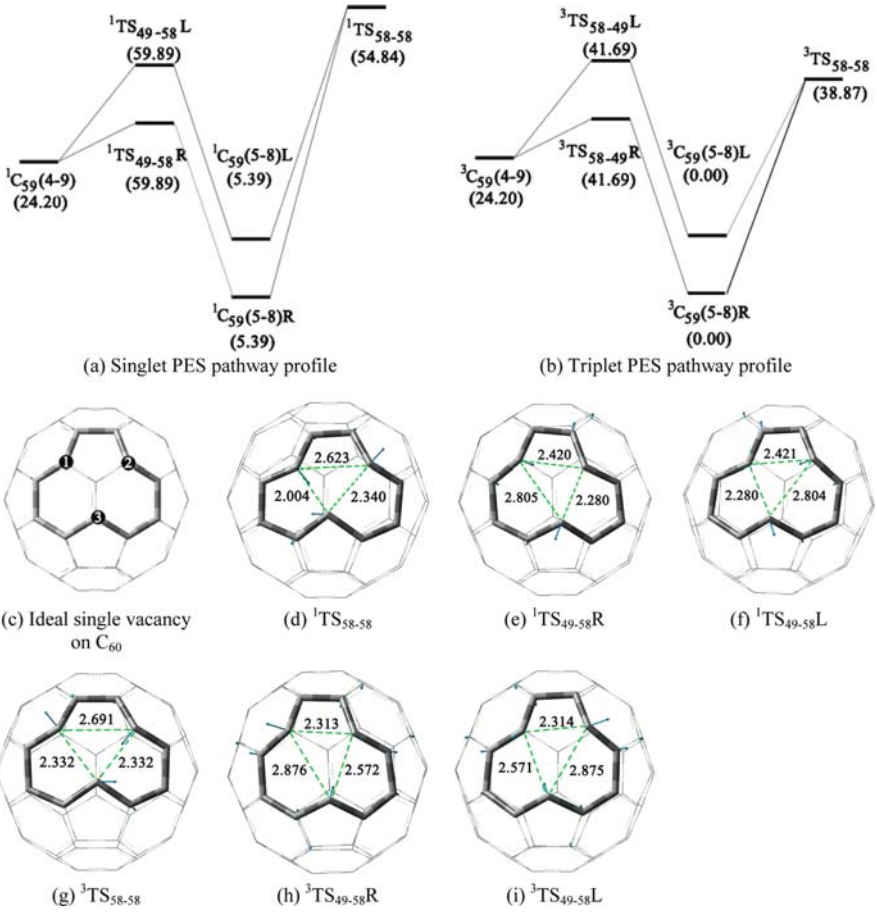


Figure 16-5. (**a**) and (**b**) are the isomerization pathway profiles of the single-vacancy defected C_{60} on the singlet and triplet potential energy surfaces, respectively (energies are in kcal/mol). (**c**) Ideal single vacancy on C_{60}; (**d**)−(**i**) are structures of the transition states (numbers are the atom distances in Å and the blue arrows represent the imaginary vibrational modes). In (**f**) and (**i**), L denotes the bond formation between C1 and C3, thus leading to a pentagon on the left. In (**e**) and (**h**), R denotes the bond formation between C2 and C3, thus leading to a pentagon on the right

Table 16-2. The vertical electron affinity (VEA), the adiabatic electron affinity (AEA), the vertical ionization potential (VIP), and the adiabatic ionization potential (AIP) of C_{59}(4-9), C_{59}(5-8), C_{69}(4-9), and C_{69}(5-8). All energies are in eV

Cluster	VEA	AEA	VIP	AIP
C_{59}(4-9)	3.42	3.32	7.36	7.29
C_{59}(5-8)	3.68	3.34	6.92	6.79
C_{69}(4-9)	3.66	3.55	7.35	7.30
C_{69}(5-8)	3.91	3.61	7.06	6.95

16.5.2. Vacancy-Defected SWCNTs

16.5.2.1. Vacancy-Defected (5,5) and (10,0) SWCNTs

Vacancies on carbon nanotubes have also been studied recently [1, 64–72, 94, 165–167, 169–175], including the effect of vacancies on the conductance [65, 66, 171], the mechanical properties [173, 174], and the electronic properties [1, 94, 172] of carbon nanotubes. During the process of vacancy creation on carbon nanotubes, energetic electrons produce mostly single vacancies, whereas heavy ion irradiation produces mostly multi-vacancies [165–167]. Removing one or more carbon atoms from a carbon nanotube first produces ideal (but unstable) vacancies (see Figure 16-6a and d). Ajayan et al. have shown that carbon nanotube will respond to the loss of carbon atoms by surface reconstruction, resulting in vacancy related point defects [64]. The nature of single vacancies and their related point defects has been studied systematically by Lu et al. with tight-binding method [70]. Double and triple vacancies and related defects have been studied by Mielke et al. [173] and Sammalkorpi et al. [174]. In these studies, however, the ground states of the defected nanotubes are not mentioned and the structural information has some inconsistencies. For example, Lu and Pan reported that the symmetric 5-1DB (one pentagon and one dangling bond) defects (Figure 16-6c) do not exist for the armchair-type SWCNT [70]. Even with the symmetric 5-1DB defect geometry as the initial guess, the asymmetric 5-1DB defect (Figure 16-6b) was obtained easily [70]. However, the symmetric 5-1DB defect on (5,5) SWCNT was shown to be stable by Mielke et al. [173]. To resolve such discrepancy, we investigated the electronic properties of single- and double-vacancy defected SWCNTs with density functional methods. (5,5) and (10,0) SWCNTs were chosen to represent typical armchair and zigzag SWCNTs, respectively. BPW91 with the 6-31G basis set are employed in geometry optimization and property prediction. The SWCNTs are modeled by imposing periodic boundary conditions (PBCs). The unit cells contain about 100 and 120 carbon atoms for (5,5) and (10,0) SWCNTs, respectively. The integrations of **k** space are achieved by using the default numbers of **k** points, 27 and 26 for (5,5) and (10,0) SWCNTs, respectively.

Perfect (5,5) and (10,0) SWCNTs are studied for comparison with defect SWCNTs. Removal of a single atom yields a dodecagon on both (5,5) and (10,0) SWCNTs (Figures 16-6a and 16-7a). After the surface reconstruction, a pentagon and a nonagon (contains an unsaturated carbon atom) appear in two different ways: one is asymmetric (Figures 16-6b and 16-7b) and the other is symmetric (Figures 16-6c and 16-7c) [174]. There are also two ways to lose two adjacent carbon atoms on both (5,5) and (10,0) SWCNTs, resulting in tetradecagons (Figures 16-6d, f, 16-7d, and f). After the surface reconstruction, two pentagons and one octagon appear in the symmetric (Figures 16-6e and 16-7e) and asymmetric (Figures 16-6g and 16-7g) patterns. No dangling bond is present for the double-vacancy defects. The energies of these isomers are shown in Table 16-3.

The (5,5) SWCNT 1 asym. isomer (Figure 16-6b) is more stable than the (5,5) SWCNT 1 sym. isomer (Figure 16-6c). The ground state of the (5,5) SWCNT 1 asym. isomer is singlet, while the ground state of the (5,5) SWCNT 1 sym. isomer

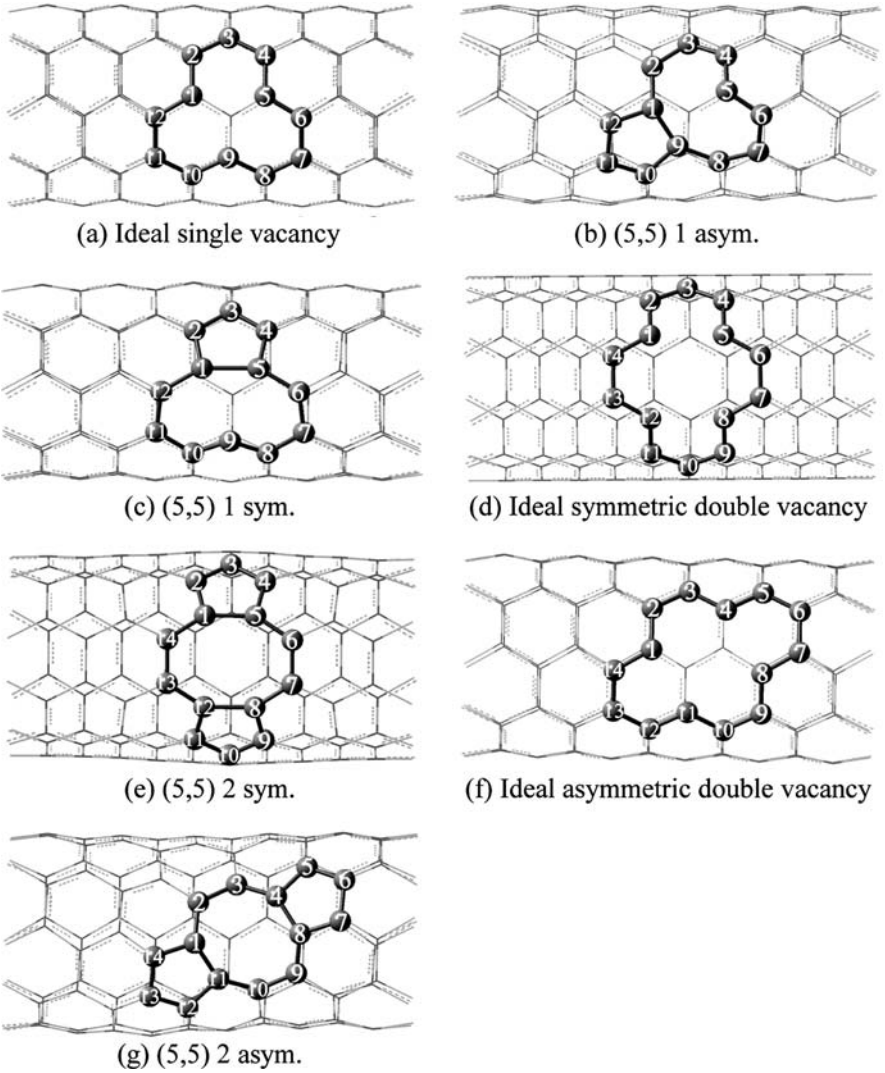


Figure 16-6. Structures of the ideal single and double vacancies and related defects on the (5,5) SWCNT

is triplet. The bridging bonds of the pentagon and the nonagon in the (5,5) SWCNT 1 sym. isomers are very weak: 1.846 and 1.871 Å in length for the singlet and triplet cases, respectively. However, for the asymmetric isomers, the bridges are typical C−C single bonds, with bond distances 1.579 and 1.565 Å for the singlet and triplet cases, respectively. These results agree with the predictions of Mielke et al. [173] very well: single-vacancy defected (5,5) SWCNT does have the symmetric 5-1DB defect, which was not found by Lu et al. with tight-binding method [70].

A finite model of the (5,5) SWCNT 1 sym. isomer, which has the same number of carbon atoms as in the PBC model and whose two open ends are capped by hydrogen

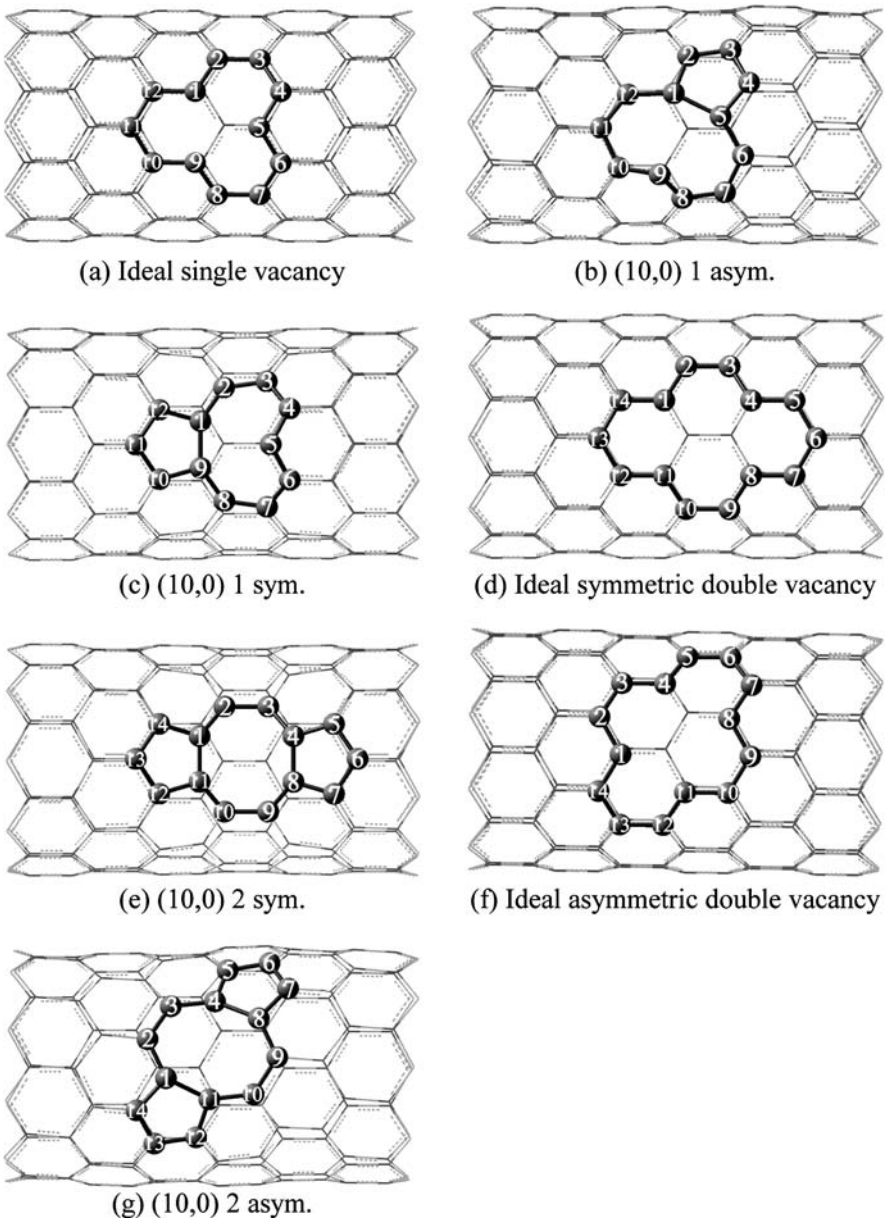


Figure 16-7. Structures of the ideal single and double vacancies and related defects on the (10,0) SWCNT

atoms, was studied for comparison. In this finite model, the bridging bonds of the pentagon and the nonagon are 1.639 and 1.649 Å in length for the singlet and triplet cases, respectively, which are shorter than their counterparts in the PBC model. This is simply due to the relaxation of the constraint at the two open ends of the finite

Table 16-3. Energies calculated at the BPW91/6-31G level of theory for all the related defects of the single- and double-vacancy defected (5,5) and (10,0) SWCNTs and the perfect (5,5) and (10,0) SWCNTs

	E_{total} (Hartree)		E_{sa}[c] (kcal/mol)		ΔE[d] (kcal/mol)	
System	S (R/U)[a]	T (U)[a]	S (R/U)	T (U)	S(R/U)	T (U)
(5,5) SWCNT	−3810.0481	–	169.20	–	–	–
(5,5) SWCNT 1 asym.	−3771.7310	−3771.7274 (2.02)[b]	167.85	167.83	1.35	1.37
(5,5) SWCNT 1 sym.	−3771.6790	−3771.6944 (2.02)	167.52	167.62	1.68	1.58
(5,5) SWCNT 2 asym.	−3733.6917	−3733.6744 (2.02)	168.23	168.12	0.97	1.08
(5,5) SWCNT 2 sym.	−3733.5905	−3733.5735 (2.01)	167.58	167.48	1.62	1.72
(10,0) SWCNT	−4572.2690	–	170.31	–	–	–
(10,0) SWCNT 1 asym.	−4533.9033	−4533.9102 (2.02)	168.94	168.98	1.37	1.33
(10,0) SWCNT 1 sym.	−4533.9535	−4533.9402 (2.02)	169.21	169.14	1.10	1.17
(10,0) SWCNT 2 asym.	−4495.8388	−4495.8144 (2.00)	169.13	169.00	1.18	1.31
(10,0) SWCNT 2 sym.	−4495.9168	−4495.9041 (2.00)	169.55	169.48	0.76	0.83

[a] R: Spin-restricted; U: Spin-unrestricted; S: Singlet; T: Triplet.
[b] The values in the parentheses are the $\langle S^2 \rangle$ values.
[c] The stability energy per atom, defined as E_{ts}/n, where E_{ts} is the energy difference between the isolated carbon atoms and the cluster and n is the total number of carbon atoms in the cluster.
[d] $\Delta E^f = E_{sa}$ (perfect SWCNT) − E_{sa} (defected SWCNT).

model. The energy of the triplet isomer is 14.10 kcal/mol lower than that of the singlet isomer, which agrees with the results of the PBC model. Thus, the existence of the symmetric 5-1DB defect on single-vacancy defected (5,5) SWCNT is without a doubt.

The (5,5) SWCNT 2 asym. isomer (Figure 16-6g) is more stable than its symmetric counterpart. The energy of the singlet state is 10.86 kcal/mol lower than that of the triplet state. The ground state of the (5,5) SWCNT 2 sym. isomers is also singlet, whose energy is 10.67 kcal/mol lower than that of the triplet state. The bond lengths of the bridging bonds of the (5,5) SWCNT 2 sym. isomer are 1.687 Å, longer than those of the (5,5) SWCNT 2 asym. isomer (1.534 Å). The removal of one or two carbon atoms from perfect (5,5) SWCNT decreases the cohesive energy per carbon atom by 1.35, 1.58, 0.97, and 1.62 kcal/mol for the (5,5) SWCNT 1 asym., 1 sym., 2 asym., and 2 sym. systems, respectively. Apparently, the (5,5) SWCNT 2 asym. isomer is the most stable one.

According to the FMO analysis (Figure 16-8A and B) of the singlet (5,5) SWCNT 1 asym. and 2 asym. systems, we can see that vacancy defects severely destruct the π conjugated system of (5,5) SWCNTs. Vacancy defects also create localized electronic states, which are clearly shown by the HOMO and the LUMO of the (5,5) SWCNT 1 asym. system around the nonagon. However, for the (5,5) SWCNT 2 asym. system, the HOMO and the LUMO are extended π bonds around the octagon and show no obvious localization.

The (10,0) SWCNT 1 sym. isomer (Figure 16-7c) is more stable than its asymmetric counterpart (Figure 16-7b). The ground state of the (10,0) SWCNT 1 sym. isomer is singlet, 8.34 kcal/mol lower in energy than the triplet state. The ground

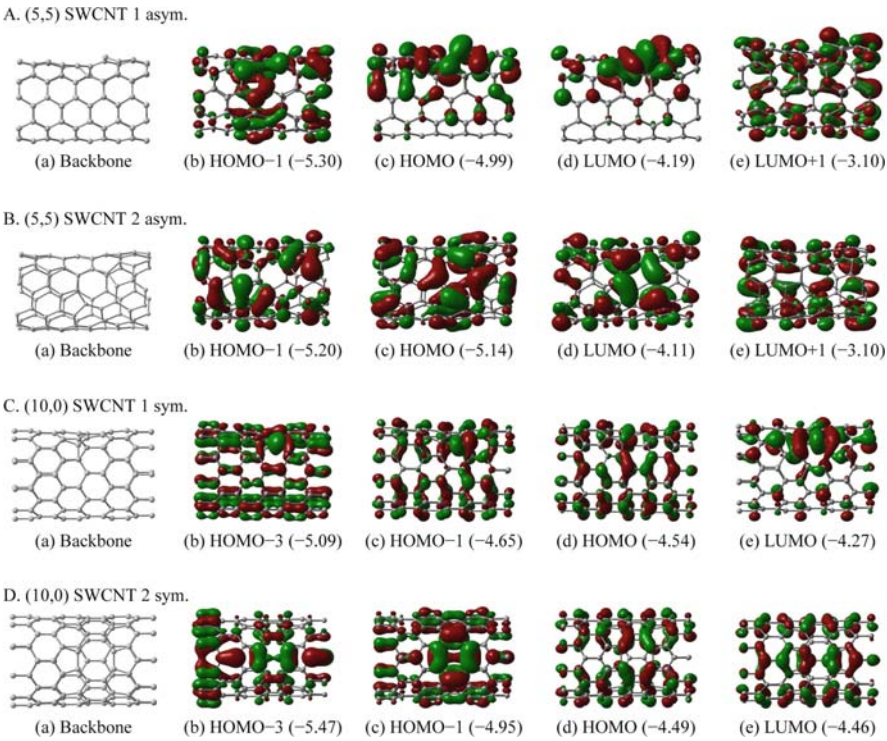


Figure 16-8. Frontier molecular orbitals of (A) the (5,5) SWCNT 1 asym. cluster, (B) the (5,5) SWCNT 2 asym. cluster, (C) the (10,0) SWCNT 1 sym. cluster, and (D) the (10,0) SWCNT 2 sym. cluster. HOMO−n (p) is the nth molecular orbital below the HOMO with orbital energy p eV. LUMO+m (q) is the mth molecular orbital above the LUMO with orbital energy q eV

state of the (10,0) SWCNT 1 asym. isomer is triplet, 4.33 kcal/mol more stable than the singlet state. The bridging bonds of the symmetric isomers are 1.560 and 1.530 Å in length for the singlet and triplet states, respectively. However, the bridging bonds for the asymmetric isomers are much longer, with bond distances of 1.761 and 1.751 Å for the singlet and triplet states, respectively.

The (10,0) SWCNT 2 sym. isomer is more stable than its asymmetric counterpart. The energies of the symmetric and asymmetric isomers in their singlet states are 8.16 and 15.31 kcal/mol lower than their triplet states, respectively. The bridging bonds for symmetric and asymmetric isomers are around 1.511 and 1.620 Å in length, respectively. The removal of one or two carbon atoms from perfect (10,0) SWCNT also decreases the cohesive energy per carbon atom by 1.37, 1.10, 1.18, and 0.76 kcal/mol for the (10,0) SWCNT 1 asym., 1 sym., 2 asym., and 2 sym. isomers, respectively. Among them, the (10,0) SWCNT 2 sym. isomer is the most stable one. According to the FMO analysis of these two systems (Figure 16-8C and D), we can see that vacancy defects also severely destruct the π conjugated system of (10,0) SWCNT. The LUMO of the (10,0) SWCNT 1 sym. system obviously has

some major contributions from the defect site. Somehow, the localization of the electronic state is not as obvious as that on (5,5) SWCNT.

16.5.2.2. Vacancy-Defected (5,5) SWCNT Clip

Removal one carbon atom from (5,5) SWCNT clip $C_{200}H_{20}$ results in a vacancy-defected SWCNT clip $C_{199}H_{20}$ with the 5-1DB defect. The existence of the 5-1DB defects may provide a possible new way to functionalize the sidewall of the SWCNT. The semiempirical MNDO-PM3 method was initially employed to optimize the geometries of $C_{200}H_{20}$ and $C_{199}H_{20}$, and the geometries were then fully optimized within BPW91/6-31G. The geometric effect due to different methods on the structure is negligible [1, 94].

The structure, DOS, and LDOS of $C_{199}H_{20}$ are shown in Figures 16-9a and b. The removal of one carbon atom from the SWCNT relaxes the geometric constrain in the curved sidewall of the SWCNT and stabilizes the HOMO and the LUMO, as indicated in Figure 16-9. The HOMO-LUMO gap of $C_{199}H_{20}$ (0.39 eV) is slightly smaller than that of $C_{200}H_{20}$ (0.41 eV). The LDOS of different regions in $C_{200}H_{20}$ corresponding to the vacancy in $C_{199}H_{20}$ are plotted to better understand the effect of creating vacancy defect to the electronic structure of the SWCNT. The contribution to the HOMO and the LUMO from the region of the vacancy defect

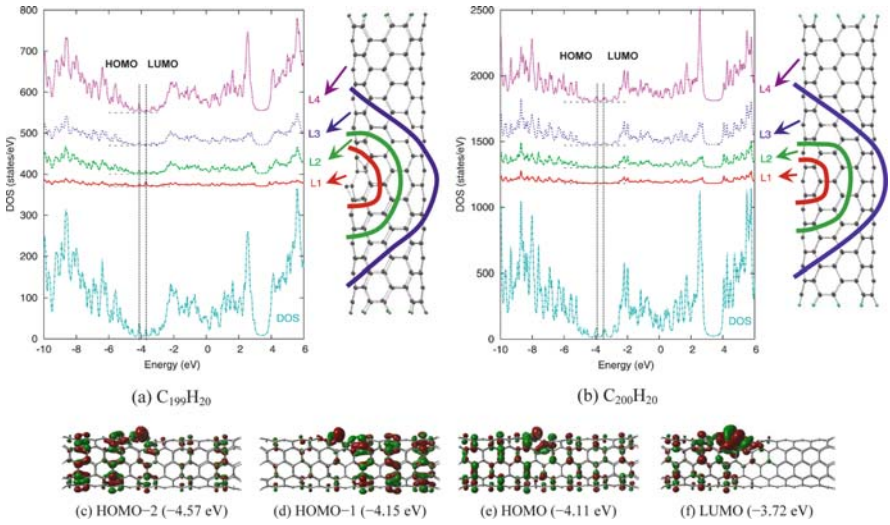


Figure 16-9. The density of states and local density of states of (**a**) the open-end vacancy-defected (5,5) SWCNT segment $C_{199}H_{20}$ and (**b**) $C_{200}H_{20}$. (**c**)−(**f**) are the frontier molecular orbitals of the open-end vacancy-defected (5,5) SWCNT segment $C_{199}H_{20}$. HOMO is the highest occupied molecular orbital, and LUMO is the lowest unoccupied molecular orbital with orbital energy −3.72 eV. HOMO−*n* (*p* eV) is the *n*th orbital below the HOMO with orbital energy *p* eV. L1−L4 are the local density of states for each specified group of atoms as outlined on the structures. This figure is adopted from Ref. [94]

in $C_{199}H_{20}$ (L1 in Figure 16-9a) is much stronger than that of the corresponding region in $C_{200}H_{20}$. There are localized states at the vacancy defect region in the HOMO and the LUMO. The π MOs in $C_{199}H_{20}$ is destructed to a large degree by the vacancy defect with respect to the MOs of $C_{200}H_{20}$. The vacancy defect region divides $C_{199}H_{20}$ into two parts at the middle nonagon through the direction perpendicular to the SWCNT axis; this division is clearly manifested by the occupied MOs from the HOMO-2 to the HOMO, especially by the HOMO-2 and the HOMO-1. The HOMO consists of lone-pair electrons from the dangling carbon atom bonding to two carbon atoms and the π bonds of the other carbon atoms of the nonagon. The LUMO has dominant contributions from the nonagon especially from the dangling carbon atom. However, only half of the SWCNT contributes to the LUMO. The MOs of $C_{199}H_{20}$ indicate that the vacancy defect region will be the active center in chemical reactions.

16.6. DOPED SWCNTS

The vacancy defects in both fullerene and SWCNT create new active centers. Another way to create active center is replacing one carbon atom with another atom such as B, N, or a metal atom, resulting in HSWCNT.

16.6.1. B- and N-Doped SWCNTs

The doping of B or N strongly modifies the electronic structure of SWCNT particularly near the Fermi level [73–77, 176], and enhances the non-linear optical coefficients [177] and the sensitivity and selectivity of carbon nanotube, thus broadening its potential application as chemical sensor [75]. Because of the electron deficient nature of B, there are prominent acceptor-like peaks close to the Fermi level of B-doped carbon nanotube [73], thus significantly enhancing the conductivity of carbon nanotube. Like the fullerene counterpart, the electron deficiency of B-doped carbon nanotube makes such carbon nanotube useful in hydrogen storage [178]. Structurely, the doping of B takes para position in a hexagon in carbon nanotube [73]. On the other hand, the doping of N in carbon nanotube adopts pyridine-like structure and brings donor-like peaks close to the Fermi level because of the electron abundance in N [176]. In different situations, N-doped carbon nanotube can be either n-type or p-type (with vacancy) device [76]. However, the mechanism of formation [179], detailed structure [73–77], and applications of such doped carbon nanotubes [94, 177] await for further investigations.

16.6.2. Ni-, Pd-, and Sn-Doped SWCNTs

Studies on metal-doped carbon nanotubes are rather sparse [1, 180, 181]. For the main group elements (such as B, N, and O) whose electronic structures and sizes are similar to those of carbon, their dopings to carbon nanotubes do not greatly distort the structures of carbon nanotubes. However, due to their large atomic radii

and partially occupied *d* orbitals, transition metal atoms have different bonding to carbon from those of the main group elements. For example, a carbon atom at the middle of the clip can be replaced with a metal atom from either outside (exo-doped) or inside (endo-doped) of the tube. Hence, two questions naturally arise: which one of the exo- and endo-doped conformations is more stable for metal-doped HSWCNT and what are the electronic structures and reactivities of these doping sites? To answer these questions, Pd-, Ni-, and Sn-doped HSWCNTs were studied with quantum mechanic methods. A (5,5) SWCNT with 70 carbon atoms saturated with 20 hydrogen atoms at two ends was chosen for the modeling.

Geometry optimizations and frequency predictions on these HSWCNTs were carried out with B3LYP/LANL2MB [182]. The energies and electronic properties were refined with B3LYP/LANL2DZ [182]. The model and optimized structures for these HSWCNTs are shown in Figure 16-10. Major bond distances and relative energies of these HSWCNTs are listed in Table 16-4. The doping metal atom essentially bonds to three carbon atoms in both endo- and exo-doped HSWCNTs. All exo-doped HSWCNTs are more stable than their corresponding endo-doped HSWCNTs. All bond distances between the doping metal atom and the bonded carbon atoms are much longer than any C−C bond, thus resulting in much distorted geometry around the doping site and different electronic structures for endo- and exo-doped HSWCNTs. Except for the Ni- and Pd-endo-doped HSWCNTs, the doping metal atoms have large amount of positive charges. The doping metal atoms are always more positively charged in the endo-doped HSWCNTs than in the corresponding exo-doped HSWCNTs.

The difference in electronic structures between the endo- and exo-doped HSWCNTs and among different HSWCNTs is illustrated by the HOMO and the LUMO of these HSWCNTs, shown in Figure 16-11. The metal atoms in all these HSWCNTs serve as both electron acceptors and donors because of their partially occupied *d* orbitals, except for Sn-doped HSWCNTs.

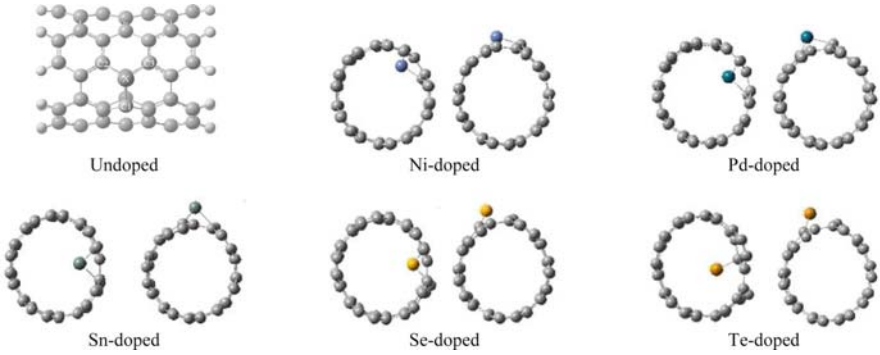


Figure 16-10. The structures of undoped (5,5) SWCNT and Ni-, Pd-, Sn-, Se-, and Te-doped (5,5) HSWCNTs

Table 16-4. Data for metal-doped (5,5) HSWCNTs. The doping atom and its neighboring three carbon atoms are donoted by X, C1, C2 and C3 as in Figure 16-11, where the latter two carbon atoms are symmetrically equivalent to each other. Energy is in eV; distance is in Å

X	Conformation	Relative energy	X−C1 distance	X−C2(C3) distance	Partial charge on X
Ni	exo	0.00	1.905	1.851	0.310
	endo	1.63	1.791	1.847	−0.004
Pd	exo	0.00	2.066	2.009	0.195
	endo	2.55	1.937	2.022	−0.104
Sn	exo	0.00	2.184	2.145	0.643
	endo	4.29	2.112	2.111	0.501
Se	exo	0.00	2.121	2.032	0.797
	endo	3.77	2.083	1.988	0.578
Te	exo	0.00	2.268	2.163	0.987
	endo	4.83	3.013	2.260	0.496

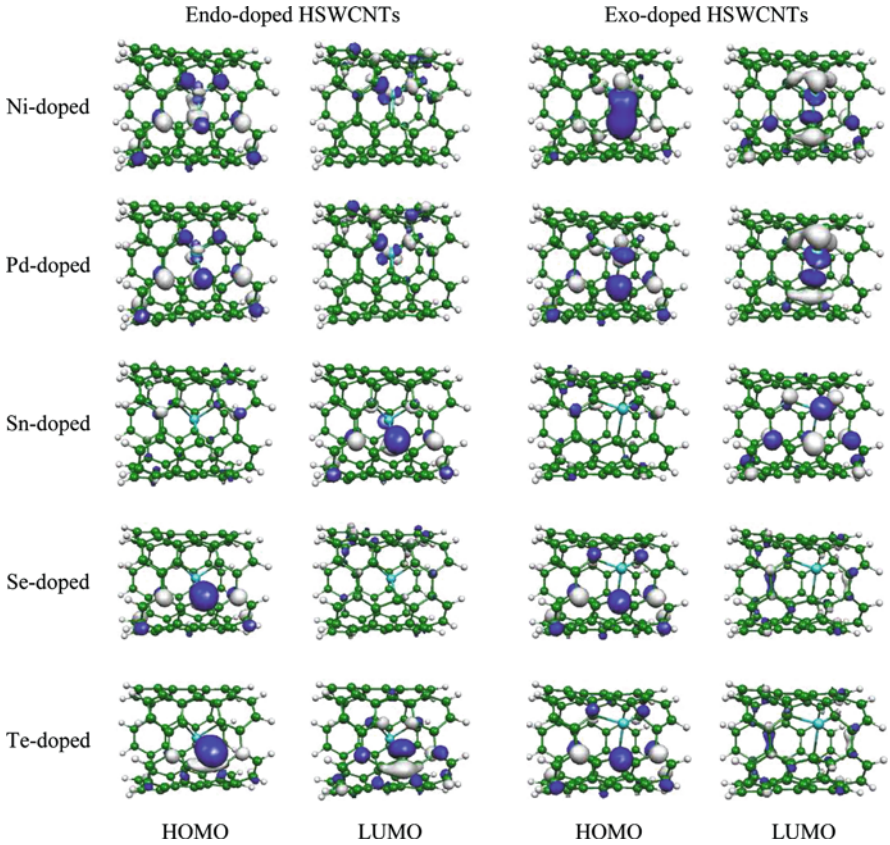


Figure 16-11. The highest occupied molecular orbitals (HOMOs) and the lowest unoccupied molecular orbitals (LUMOs) of metal-doped (5,5) HSWCNTs

16.6.3. Chalcogen Se- and Te-Doped SWCNTs

Chalcogen Se- and Te-doped HSWCNTs were studied by the same method. The structures, the HOMO, and the LUMO of these HSWCNTs are shown in Figures 16-10 and 16-11. Relative energies, bond distances, and charges for the doping atoms are listed in Table 16-4. Se and Te are positively charged in both the exo- and endo-doped HSWCNTs. The HOMOs in all these HSWCNTs are localized around the doping site. The LUMOs of Se-doped HSWCNTs and Te-exo-doped HSWCNT are delocalized along the tube.

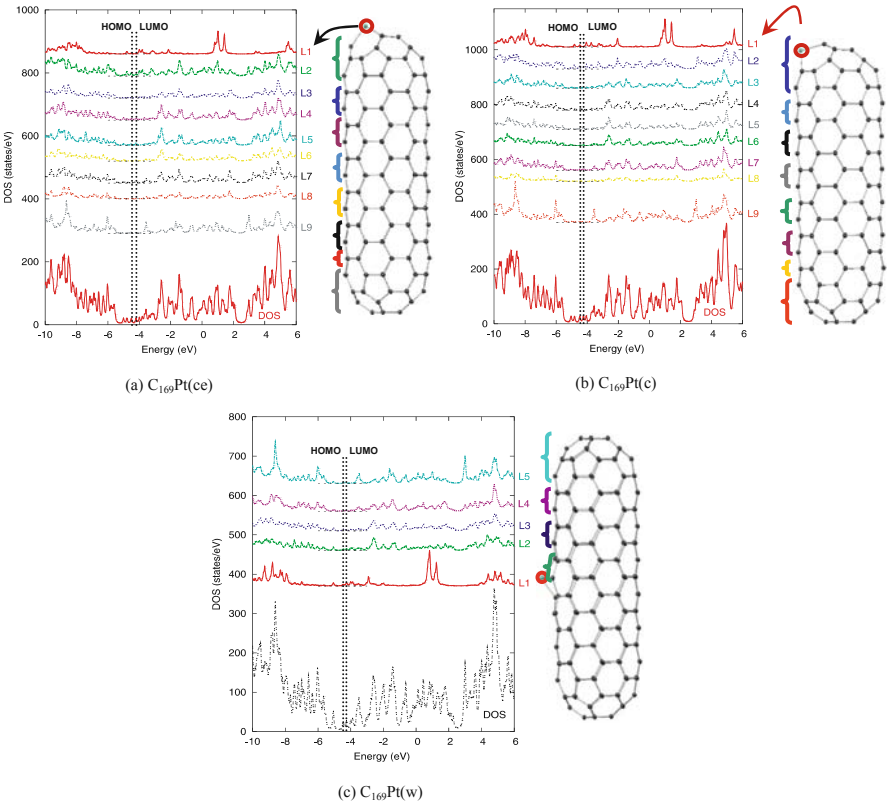
16.6.4. Pt-Doped SWCNTs

From above, we conclude that the exo-doped HSWCNT is more stable than the endo-doped HSWCNT and there exists localized electronic states at the doping site. Such conclusion could be also true for HSWCNT rods if the doping site is at the middle of the sidewall [1, 180]. However, the hemispherical caps of SWCNT rod have different curvature from that of the sidewall, thus can have different electronic structure and reactivity [10, 11, 17, 138]. To appreciate such difference in the electronic structure and reactivity, Pt-doped HSWCNT rods originated from (5,5) SWCNT were studied with BPW91/LANL2DZ.

For (5,5) SWCNT rod C_{170}, the NBO partial charge analysis indicates that the five atoms connecting to the top pentagon of the cap have the largest negative charges and the atoms of the next layer have the largest positive charge. Thus, the caps can be chemical reaction centers. Substitution of the carbon atoms in the cap by other elements will change the chemical selectivity and sensitivity of the SWCNT rod in catalytic reactions. Replacing one carbon atom with a Pt atom in the end pentagon, in the next layer, or on the sidewall of C_{170} results in three Pt-doped SWCNT rods: C_{169}Pt(ce), C_{169}Pt(c), and C_{169}Pt(w). Single-point calculations at the BPW91/6-31G level of theory predict C_{169}Pt(ce) is the most stable nanorod: the total energies of C_{169}Pt(c) and C_{169}Pt(w) are 0.8 and 17.9 kcal/mol higher, respectively. Evidently, the cap-doped SWCNTs are more stable than the wall-doped SWCNT, because of the relaxation of the constraint on the cap through doping. The total energy of the triplet electronic state of the Pt-doped nanorod is found to be higher than that of the singlet, *i.e.* the ground state of the Pt-doped nanorod is singlet.

Figure 16-12 displays the structures, DOS, and LDOS of the cap-end-doped C_{169}Pt(ce), the cap-doped C_{169}Pt(c), and the wall-doped C_{169}Pt(w). Their FMOs are shown in Figure 16-13. The shapes of the LDOS of different layers are similar at the FMO region, which indicates the delocalization of the FMOs of the HSWCNT rods.

The Pt−C bond distances at the cap end are 2.01 Å and the other Pt−C bond distances are 1.96 Å. Clearly, the Pt atom points outward along the translation direction of the SWCNT rod. The distortion of the SWCNT rod due to the Pt-doping is localized on the pentagons and hexagons around Pt. The overall DOS of C_{169}Pt(ce)



(a) $C_{169}Pt(ce)$ (b) $C_{169}Pt(c)$

(c) $C_{169}Pt(w)$

Figure 16-12. (**a**) Density of states and local density of states for the Pt cap-end-doped nanorod $C_{169}Pt(ce)$ with C_s symmetry. The HOMO has orbital energy −4.51 eV, and the LUMO has orbital energy −4.21 eV. (**b**) Density of states and local density of states for the Pt cap-doped nanorod $C_{169}Pt(c)$ with C_s symmetry. The HOMO has orbital energy −4.48 eV, and the LUMO has orbital energy −4.25 eV. (**c**) Density of states and local density of states for the Pt wall-doped nanorod $C_{169}Pt(w)$ with C_s symmetry. The HOMO has orbital energy −4.46 eV, and the LUMO has orbital energy −4.26 eV. L1−L9 are the local density of states for each specified layer of atoms as marked on the structures. This figure is adopted from Ref. [180]

is similar to that of C_{170} with D_{5h} symmetry. The doping of Pt produces more peaks in the DOS in the FMO region due to the introduction of the Pt 5*d* orbitals and the induced electronic structure change in the doped cap.

The HOMO of $C_{169}Pt(ce)$ is similar to that of C_{170}, except for the significant contributions from Pt and its neighboring carbon atoms in $C_{169}Pt(ce)$. The effect of Pt on the electronic structure of $C_{169}Pt(ce)$ is also reflected in several other frontier occupied MOs [1, 180]. The HOMO−1 has some major contributions from the Pt 5*d* orbitals, which form *d-p* π bonds with the carbon atoms in the next layer. The geometric distortion in $C_{169}Pt(ce)$ induces single and double CC bond alteration around Pt, which is reflected in the strong π bonding around the doped cap in the

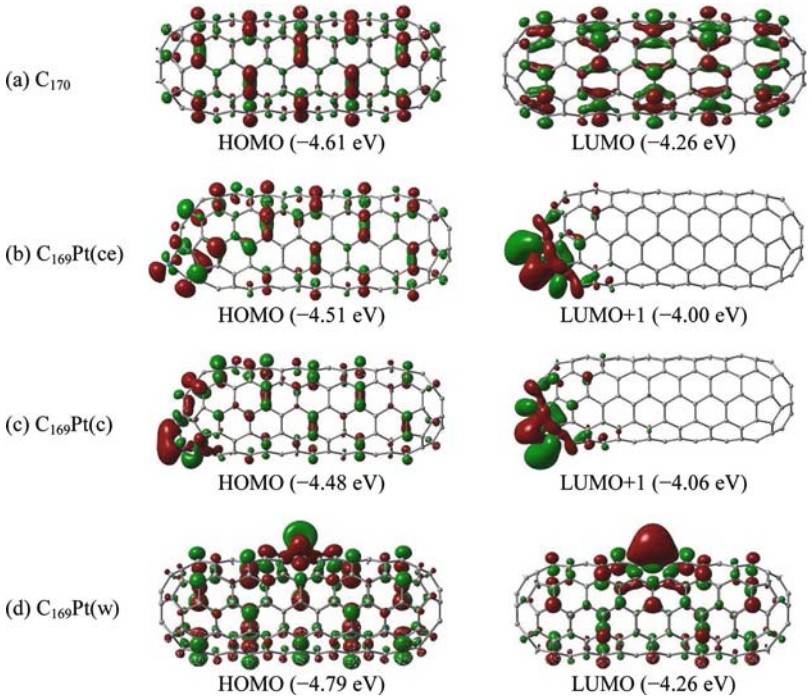


Figure 16-13. Relevant frontier orbitals of the Pt wall-doped nanorods: (**a**) C_{170} with D_{5h} symmetry, (**b**) $C_{169}Pt(ce)$ with C_s symmetry, (**c**) $C_{169}Pt(c)$ with C_s symmetry, and (**d**) $C_{169}Pt(w)$ with C_s symmetry. LUMO+1 (q eV) is the first orbital above the LUMO with orbital energy q eV. This figure is adopted from Ref. [180]

HOMO−5 [180]. The LUMO of $C_{169}Pt(ce)$ is very similar to that of C_{170}. The next unoccupied MO, the LUMO+1 (and the LUMO+2), of $C_{169}Pt(ce)$ is mainly from the 5*d* orbitals of Pt, and the contribution to this unoccupied MOs from the un-doped cap diminishes. According to the MOs of $C_{169}Pt(ce)$, it is quite clear that the reactive center in $C_{169}Pt(ce)$ is around the location of Pt.

The Pt atom can donate electrons to electron acceptors and its empty 5*d* orbitals can accept electrons from electron donors, *e.g.* in reaction with gases like H_2, C_2H_4, CO, NH_3, NO, etc. The NBO analysis indicates that Pt transfers about 0.80 electrons (0.40 electrons from 6*s* and 0.40 electrons from 5*d*) to the nearby carbon atoms: the electronic configuration of Pt is essentially $[Xe]5d^{8.60}6s^{0.60}$. The carbon atom connecting to Pt in the second layer from the doped cap end has the largest negative charge, –0.22.

The DOS and the LDOS of each layer of atoms of $C_{169}Pt(c)$ are very similar to those of $C_{169}Pt(ce)$. The structure of $C_{169}Pt(c)$ is also similar to that of $C_{169}Pt(ce)$, except for the region around Pt. The FMOs of $C_{169}Pt(c)$ are very similar to those of $C_{169}Pt(ce)$ [180]. The noticeable difference is that Pt contributes to the HOMO of $C_{169}Pt(c)$ more than it does in $C_{169}Pt(ce)$, which enhances the reactivity of Pt in $C_{169}Pt(c)$. The electronic configuration of Pt is essentially $[Xe]5d^{8.63}6s^{0.60}$. The

partial charge of Pt is 0.77 and the carbon atom connecting to Pt in the top pentagon has the largest negative partial charge, –0.25.

The DOS of $C_{169}Pt(w)$ around the FMO region is different from those of $C_{169}Pt(ce)$ and $C_{169}Pt(c)$: the contribution to the DOS at the FMO region from Pt has noticeably increased in $C_{169}Pt(w)$. The LDOS of Pt indicates that Pt contributes significantly to the DOS at the FMO region in $C_{169}Pt(w)$. The introduction of Pt on the sidewall of the SWCNT also drastically changes the LDOS of its neighboring carbon layers, as vividly demonstrated by the comparison of the LDOS of L2 and L3 with L1 and L2 of C_{170} in Figure 16-12 and the HOMO and the LUMO in Figure 16-13. From the HOMO−1 to the LUMO+3, Pt has significant contributions to each MO [180]. The HOMO-LUMO gap (0.53 eV) of $C_{169}Pt(w)$ is larger than those of C_{170} (0.35 eV), $C_{169}Pt(ce)$ (0.30 eV), and $C_{169}Pt(c)$ (0.23 eV). In chemical reactions, Pt will serve as various catalytic centers with flexible oxidation states capable of accepting and donating electrons. The electronic configuration of Pt is essentially $[Xe]5d^{8.63}6s^{0.54}$. The partial charge of Pt is 0.83. The partial charge of the carbon atom connecting to Pt in the symmetric plane is –0.18 and the partial charges of the other two equivalent carbon atoms connecting to Pt are –0.10.

In summary, the doping of Pt in SWCNT significantly changes the local electronic structure of the HSWCNT especially around the doping site, thus creating new chemical active center.

16.6.5. Gas Adsorptions on Pt-Doped SWCNTs

Doping of Pt enhances both the electron accepting and donating capacities of the doped nanorod, changes the chemical reactivity and regioselectivity of the SWCNT, and broadens the field of the applications of the SWCNT rods in such areas as gas sensors [183]. Present studies have found that the change of structure and reactivity through the doping of Pt in the SWCNT is localized at the doping site. Adsorptions of C_2H_4 and H_2 on the Pt atom of the three Pt-doped SWCNTs reveal the different reactivity of the Pt-doped SWCNTs. The relative stability of the Pt-doped SWCNTs and the adsorption energies of C_2H_4 and H_2 on the Pt-doped SWCNTs are collected in Table 16-5. Figure 16-14 shows the bond distances at the adsorption site.

Table 16-5. The relative stabilities of the Pt-doped nanorods and the adsorption energies of C_2H_4 and H_2 on the Pt-doped nanorods. The minus sign indicates the release of the heat of formation upon the adsorption [180]. All energies are in kcal/mol

Adsorbate	Level of theory	$C_{169}Pt(ce)$	$C_{169}Pt(c)$	$C_{169}Pt(w)$
None	BPW/6-31G	0.0	0.8	17.9
	B3LYP/Lanl2mb	0.0	2.4	28.6
	B3LYP/Lanl2dz	0.0	3.3	26.3
C_2H_4	B3LYP/Lanl2mb	−22.4	−23.5	−19.2
	B3LYP/Lanl2dz	−20.6	−22.2	−14.7
H_2	B3LYP/Lanl2mb	−2.1	−2.3	−10.7
	B3LYP/Lanl2dz	−2.8	−3.0	−10.9

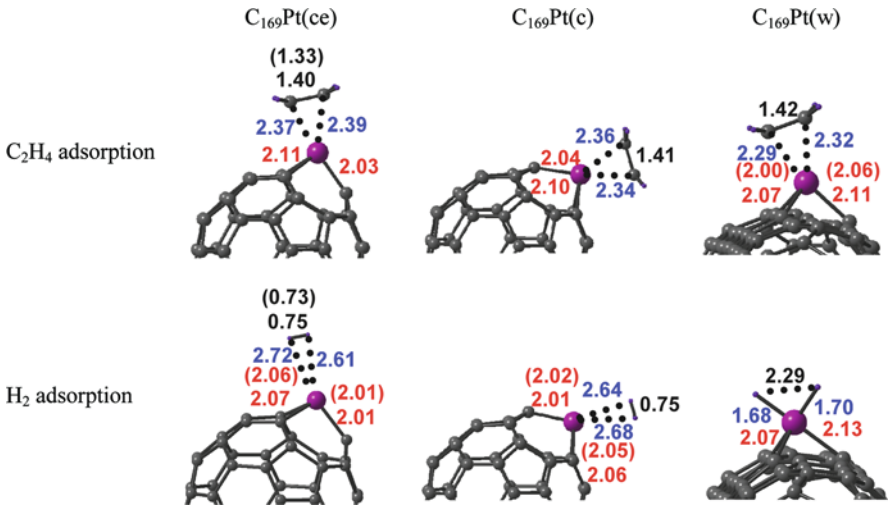


Figure 16-14. The relevant bond distances (in Å) of the adsorptions of C_2H_4 and H_2 on the Pt atom in the Pt-doped nanorods ($C_{169}Pt$). The Pt atom is in purple. The C=C bond distances in the C_2H_4 adsorption and the HH bond distances in the H_2 adsorption are in black. The PtC bond distances between Pt and the carbon atoms of C_2H_4 in the C_2H_4 adsorption and the PtH bond distances in the H_2 adsorption are in blue. The PtC bond distances between Pt and its nearest carbon atoms of the SWCNT are in red. The numbers in parentheses are the PtC bond distances of the isolated $C_{169}Pt$ without adsorption, the C=C bond distance in the isolated free C_2H_4, and the H−H bond distance of the isolated free H_2. This figure is adopted from Ref. [180]

As exemplified by the Pt−C bond distances, there is no significant structural change at the Pt-doped region for the adsorption of H_2 on the two cap-doped SWCNT rods. These two cases are physisorptions according to the H−H bond distance and the distances between H_2 and the SWCNT rod (Figure 16-14) and the adsorption energies (Table 16-5). On the other hand, the adsorption of H_2 on the Pt atom in the middle of the nanorod $C_{169}Pt(w)$ is a chemical one. Obviously, the H−H bond is broken and the two hydrogen atoms form chemical bonds with Pt with bond lengths of *ca* 1.70 Å. The distance between the two hydrogen atoms is 2.29 Å. This chemisorption releases about 10.0 kcal/mol energy, which is nearly five times of those energies (about 2.0–3.0 kcal/mol) released by the adsorption of H_2 on the Pt atom at the end cape of the SWCNT rods. The interaction between H_2 and Pt in the two physisorptions is mainly the electron transfer from the bonding orbital of H_2 to the empty 5*d* orbital of Pt [180]. Though the two hydrogen atoms in the adsorption on $C_{169}Pt(w)$ are separated, the interaction between these two hydrogen atoms remains strong, as revealed by the MO overlaps between them [180]. The strong bonding between Pt and H indicates the possibility of more H_2 can be adsorpbed on Pt, which opens a new path in search of hydrogen storage material.

The adsorption of C_2H_4 on the Pt atom in the three Pt-doped nanorods is physisorption with C−Pt distances *ca* 2.30 Å. As the adsorption site changes from the end-capes of $C_{169}Pt(ce)$ and $C_{169}Pt(c)$ to the middle of the sidewall of

$C_{169}Pt(w)$, the CC bond distance in C_2H_4 gets longer, and Pt−C bond distances between C_2H_4 and Pt get shorter, perhaps indicating the strength of the interaction between C_2H_4 and the Pt-doped SWCNTs in this ascending order: $C_{169}Pt(ce) < C_{169}Pt(c) < C_{169}Pt(w)$. However, this conclusion based on structure analysis alone does not agree with the data in Table 16-5: the adsorption energy of C_2H_4 is the smallest on $C_{169}Pt(w)$ and the largest on $C_{169}Pt(c)$. The trend of the adsorption strengths of C_2H_4 on the Pt-doped SWCNTs is the compromise of the weakening of the C=C double bond in C_2H_4, the electrostatic attraction between the two carbon atoms of C_2H_4 and Pt, and the repulsion between the C_2H_4 and the SWCNT rod. It is also interesting to note that the geometries of the adsorptions of C_2H_4 on the Pt atom at the end-cap of the Pt-doped SWCNTs are very similar to those on the Pt-doped fullerene, $C_{59}Pt$ [88], which should possess similar adsorption strengths.

Overall, the adsorptions of H_2 and C_2H_4 get stronger as the adsorption site changes from the hemispheric cap to the sidewall of the SWCNT rod, as manifested by the relevant MOs [180] and the adsorption energies, reflecting different chemical activity on different sites.

16.7. CHEMICAL REACTIONS OF VACANCY-DEFECTED SWCNTs

MO and LDOS analyses reveal that the vacancy defect on SWCNT is chemical active. Particularly, the carbon atom with dangling bond of the 5-1DB defect is the active center. The modelings of the reactions of vacancy-defected (5,5) SWCNT clip with NO [94] and O_3 [184] provide strong evidence for the reactivity of the vacancy defect. Recent modeling of H_2 adsorbed at the vacancy in vacancy-defected (12,0) SWCNT revealed that after overcoming a potential barrier, H_2 is strongly adsorbed on the carbon atom with dangling bond [185].

16.7.1. Computational Details and Model Selection

The electronic state of the 5-1DB vacancy is localized in vacancy-defected (5,5) SWCNT according to the MO and LDOS analyses. Such localization enables the treatment of the vacancy within the ONIOM model [94, 118, 184]. To validate the modeling selection, a single-point calculation of the nonagon of the 5-1DB defect capped by hydrogen atoms, C_9H_8 (shown in Figure 16-15b), was carried out at the B3LYP/6-31G level of theory. Figure 16-15 clearly shows that the HOMO and the LUMO, the most important MOs in chemical reactions according to Fukui frontier orbital theory [186], of $C_{199}H_{20}$ are very close to those of C_9H_8. Hence, the C_9H_8 model can be used to represent most of the chemical properties of the 5-1DB defect on the SWCNT [94].

A two-layered ONIOM model was used for the system. The nonagon of the 5-1DB defect and NO molecules were included in the high layer, which was treated at the B3LYP/6-31G(d) level of theory [94]. All other carbon atoms were set in the low layer, treated by UFF. The spin-unrestricted DFT method was applied to open-shell species. Partial charges shown in Figure 16-15b indicate that the carbon atom with dangling bond is the active center.

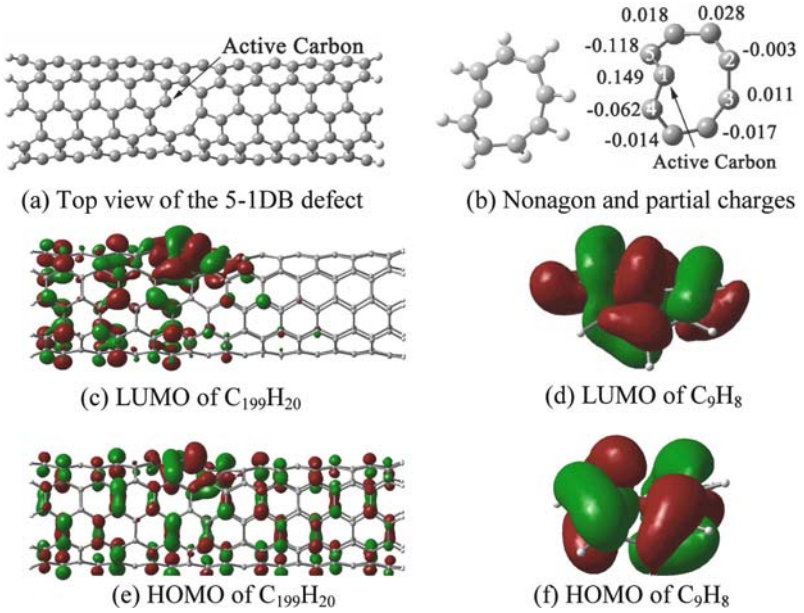


Figure 16-15. (**a**) The *top view* of the 5-1DB defect on the (5,5) SWCNT, where the nonagon is chosen as the high layer and the other carbon atoms are chosen as the low layer in a two-layered ONIOM model; (**b**) The nonagon (with partial charges) of the 5-1DB defect capped by hydrogen atoms; (**c**) The LUMO of$C_{199}H_{20}$; (**d**) The LUMO of C_9H_8; (**e**) The HOMO of $C_{199}H_{20}$; (**f**) The HOMO of C_9H_8. The figure is adopted from Ref. [94]

16.7.2. Chemical Reaction of NO with Vacancy-Defected SWCNT

The reaction of NO with $C_{199}H_{20}$ proceeds in two steps: attacking of the first NO to the tube network and extracting of the oxygen atom of the first attacking NO by the second attacking NO. The major structures of the active site in the reaction and relative energies of the stationary points on the reaction path are shown in Figure 16-16.

The active carbon atom C1 has the smallest steric hindrance and is the most chemically reactive center. The partial charges of NO are distributed as N(+0.181)−O(−0.181), whereas the active carbon atom has a +0.149 partial charge. When NO is far away from C1, the electrostatic effect should be the dominant factor controlling the initial reaction. A transition state (TS1) for the O-end attacking mode is found for two possible attacks with N or O attacking the active carbon atom. The bond length of C2 and C3 in TS1 is 1.55 Å, a typical single C−C bond. The distance between O and C1 is 1.70 Å. The energy barrier is only 8.6 kcal/mol, which means the initial attacking is very feasible, mainly due to the strong electrostatic attraction and molecular orbital overlap between O and C1. At this stage, the pentagon of the 5-1DB defect still exists, and there are 0.052 electrons transferred from the SWCNT to NO. From the shapes and energies of the FMOs of TS1, NO and the SWCNT with the 5-1DB defect, the orbital interaction for TS1 can be viewed

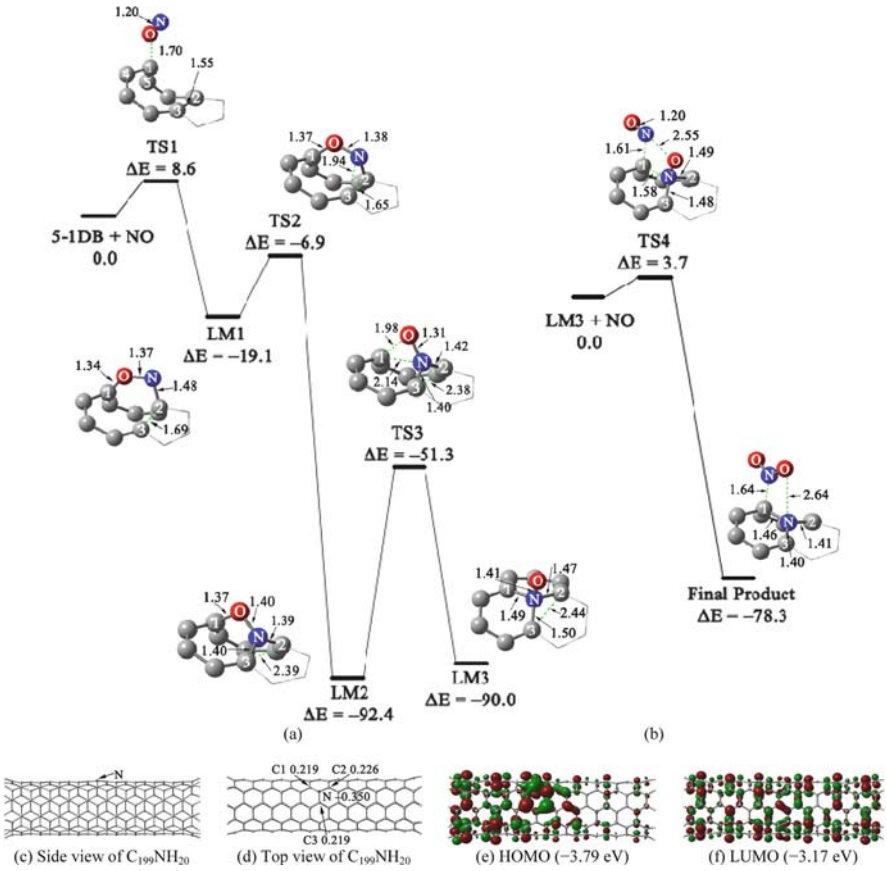


Figure 16-16. Reaction profiles and geometries of the transition states, intermediates, and finally product of the reaction of NO molecules with $C_{199}H_{20}$. The units of energy and bond length are in kcal/mol and Å, respectively. The active carbon atom is labeled as Carbon 1, and other two important carbon atoms are labeled as Carbons 2 and 3. The nitrogen atoms are in *blue*, and the oxygen atoms are in *red*. (**a**) is the reaction profile of NO with 5-1DB defected (5,5) SWCNT and (**b**) is the reaction profile of NO with LM3. (**e**) and (**f**) are the HOMO and the LUMO of $C_{199}NH_{20}$, respectively. This figure is adopted from Ref. [94]

as the single occupied molecular orbital of NO interacting with the HOMO of the SWCNT with the 5-1DB defect [94].

After overcoming the initial reaction barrier, the system reaches the first intermediate, LM1, with a bridge conformation. The oxygen atom is still chemically bounded to C1, and the nitrogen atom begins to form chemical bond with C2. When O approaches C1, N also approaches C2, with an in-phase orbital overlap. In LM1, the pentagon begins to open, and the C2−C3 bond is nearly broken with an elongated bond length 1.69 Å. The energy of LM1 is 19.1 kcal/mol lower than that of the reactants: the initial reaction is exothermic. Due to the breaking of the C2−C3 bond, C3 in LM1 becomes a new active center because it has one dangling bond.

System can easily overcome an energy barrier of 12.2 kcal/mol from LM1 reaching a transition state, TS2, which is smaller than the energy released from the initial reaction (19.1 kcal/mol). After crossing TS2, system reaches LM2 forming a hexagon of five carbon atoms and one nitrogen atom. The formation of LM2 is highly exothermic: the system releases 85.5 kcal/mol. Such a big amount of energy release is enough for the system to overcome TS3 with reaction barrier of 41.1 kcal/mol. TS3 connects to the two intermediates that have either O or N bounded to C1. The forward intermediate is LM3. In LM3, one can see that N forms single bonds with C1, C2, and C3, and also with O. The oxygen atom sticks out of the sidewall of the nanotube. The energy of LM3 is 90.0 kcal/mol lower than that of the reactants and is only 2.4 kcal/mol higher than LM2, indicating that LM3 is quite stable. In the end of the first step, the net reaction of NO with $C_{199}H_{20}$ is that NO inserts its N into the defect site with the initial attack of O toward the active carbon atom. This reaction is highly thermally feasible.

The relatively long bond distance between N and O (1.41 Å) indicates that O is not strongly bounded to N, so this N−O bond may be easily broken upon proper attack from another NO molecule in the NO excess environment. The attack of a second NO toward LM3 ($C_{199}H_{20}NO$) proceeds.

The N end of the second NO attacks C1 and O1 (the oxygen atom of the first NO) to reach TS4. The distance between C1 and N2 (the nitrogen atom of the second NO) in TS4 is 1.61 Å; the C1−N1 (the nitrogen atom of the first NO) bond elongates to 1.58 Å; and the distance between N2 and O1 is 2.55 Å. The electrostatic attraction between N2 and O1 stabilizes TS4, whereas the electrostatic repulsion between C1 and N2 counterbalances this attraction. The overall interaction of these two reactants renders the reaction barrier for TS4 to be only 3.7 kcal/mol: this reaction is very facile.

A complex of NO_2 with $C_{199}H_{20}N$ forms following TS4. In the complex, NO_2 bonds to $C_{199}H_{20}N$ through a long ionic N2−C1 bond with a bond length of 1.64 Å, and the interaction between the oxygen atom (O1) of NO_2 and the nitrogen atom (N1) of $C_{199}H_{20}N$ is very weak with a bond distance of 2.64 Å. In an experiment setting, NO_2 can be removed from the surface of the SWCNT by a flow of Ar gas [35]. The formation of the final product releases 78.3 kcal/mol energy from the reaction of the second NO with $C_{199}H_{20}NO$. The reverse reaction barrier of 82.0 kcal/mol makes the reverse reaction kinetically virtually impossible under normal conditions. In short, the second NO extracts the oxygen atom from $C_{199}H_{20}NO$, forming NO_2 and the N-substitutionally doped SWCNT through a one-step reaction.

The doped N stays slightly above the sidewall surface of the SWCNT in the N-substitutionally doped (5,5) SWCNT ($C_{199}NH_{20}$) (Figure 16-16c), because of the longer σ C−N bonds and the asymmetric sp^3 hybridization of N. The HOMO-LUMO gap is only 0.62 eV, which is much smaller than those of $C_{200}H_{20}$ and $C_{199}H_{20}$ (1.38 and 1.12 eV, respectively). Thus, the N-substitutionally doped (5,5) SWCNT has much enhanced conductivity. Substitutional doping of N destructs the conjugated π system of $C_{200}H_{20}$ and introduces localized electronic states (Figures 16-16e and 16f) like what the 5-1DB defect does to perfect SWCNT, but to a less degree.

16.7.3. Chemical Reaction of O_3 with Vacancy-Defected SWCNT

The net reaction of NO with a vacancy-defected (5,5) SWCNT ($C_{199}H_{20}$) is the insertion of N into the defect site to form an N-doped HSWCNT. Is it possible to produce O-doped HSWCNT with reaction of O_3 with vacancy-defected open-end SWCNT segment $C_{119}H_{20}$? What is the product of such reaction? Reaction of O_3 with $C_{119}H_{20}$ was studied similarly hereafter [184].

Static potential energy curves for possible reaction pathways of the reaction were explored with a 2-layered ONIOM[B3LYP/6-31G(d):AM1] model, in which the 5-1DB nonagon (Figure 16-17a) and O_3 were chosen as the high layer and the rest part was the low layer. In order to identify the most possible reaction path on the static PES, QMMD-based atom-centered density matrix propagation (ADMP) method [133–136] was employed in MD simulations for the reaction of O_3 with $C_{119}H_{20}$. The successful application of ADMP in chemical reactions [187] ensures the reliability of our MD simulation studies at 300 K with time step 0.25 fs and fictitious electronic mass 0.1 au. A 2-layered ONIOM/ADMP scheme was applied for the QMMD simulation [188]. The partition of the ONIOM model in the ADMP calculations was the same as that of the static quantum mechanical calculations, except that the UFF force field was used for the low layer to save computational time.

In the 5-1DB nonagon, C1 is the active carbon atom and is the major contributor to the HOMO and the LUMO of the vacancy-defected SWCNT [180]. The out-of-plane geometry and the large positive charge of C1 make it more reactive toward the attacking reagents. The bond lengths and the bond orders indicate that around

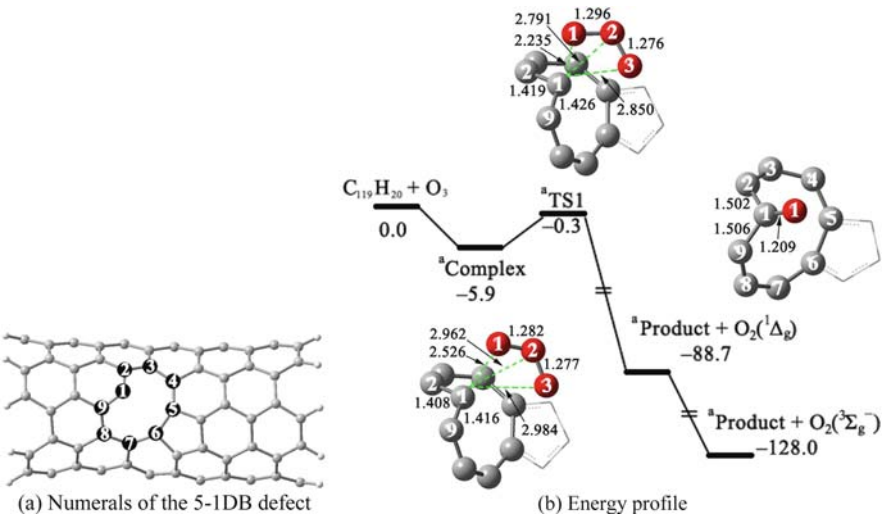


Figure 16-17. Geometries of the reaction center of the transition states, intermediates, and finally product of the reaction of O_3 with the active carbon atom of $C_{119}H_{20}$. The units of energy and bond length are in kcal/mol and Å, respectively. The oxygen atoms are in *red*. This figure is adopted from Ref. [184]

the 5-1DB defect site, the C2−C3, C4−C5, C6−C7, and C8−C9 bonds have some significant π bond characters. In fact, these four bonds have similar bond lengths to the other aromatic C=C double bonds on the sidewall of the SWCNT. Thus, the attacking of O_3 on C1 and these four bonds in the nonagon were studied as prototype reactions of O_3 with the 5-1DB defect site. The ADMP simulations will be presented at the end to verify the reaction pathway.

16.7.3.1. Reaction of O_3 with the Active Carbon Atom

The structures of the initial intermediate (^{a}Complex) and the initial attacking transition state on C1 (^{a}TS1) are shown in Figure 16-17b. In ^{a}Complex, O_3 and C1 form a tetragon. The electrostatic attraction between C1 and O1 and between C1 and O3 stabilizes ^{a}Complex [184], whereas the electrostatic repulsion between C1 and O2 counteracts with this attractive stabilization. In ^{a}TS1, O_3 and C1 form a more compact tetragon than the one in ^{a}Complex. The imaginary vibrational mode shows the shortening of the C1−O1 bond and elongating of the O1−O2 bond, indicating the attacking of O1 to C1 and the breaking of the O1−O2 bond, i.e. the dissociation of O_3. Based on the NBO analysis of ^{a}TS1, the partial charges of O1, O2, O3, and C1 are −0.21, 0.22, −0.21, and 0.14, respectively. So, there are 0.20 electrons transferred from the SWCNT to O^3, and the electrostatic interaction scheme in ^{a}TS1 is qualitatively the same as that in ^{a}Complex. Following the imaginary vibrational mode forward, C1 indeed captures O1 and the outgoing singlet O_2 ($^1\Delta g$) can be immediately quenched into a triplet O_2 ($^3\Sigma g^-$). The C1−O1 bond is very strong with a bond length of 1.209 Å, close to a typical carbonylic C=O double bond. The structure of the product (^{a}Product) is shown in Figure 16-17b. The energies of ^{a}Complex and ^{a}TS1 are 5.9 and 0.3 kcal/mol lower than the reactants, respectively. The forward reaction barrier is only 5.6 kcal/mol for ^{a}Complex, and the effective forward reaction barrier of the overall reaction is very close to zero, only −0.3 kcal/mol. The overall energy released from the reaction is −128.0 kcal/mol. Thus, the reaction of O_3 with C1 is highly exothermic and facile.

16.7.3.2. Reaction of O_3 with the C8−C9 Bond (Position 1)

The attack of O_3 on C8−C9 bond first forms an initial intermediate, ^{P1}Complex, and goes through a transition state, ^{P1}TS1 (Figure 16-18). ^{P1}Complex was reported as a π-complex in previous matrix spectroscopic and theoretical studies of ozonization of certain alkenes [189, 190]. The complex is stabilized by the interaction of the π-type HOMO and LUMO on alkenes and O_3. ^{P1}Complex is 4.3 kcal/mol more stable than the reactants, whereas ^{P1}TS1 lies 2.4 kcal/mol above the reactants. In the first intermediate ^{P1}LM1, the C8−C9 bond length is 1.663 Å, longer than those of ^{P1}Complex (1.438 Å) and ^{P1}TS1 (1.483 Å), due to the rehybridization of C8 and C9 from sp^2 to sp^3. The binding energy of ^{P1}LM1 is −27.2 kcal/mol (relative to the reactants), and the reaction of O_3 with C8−C9 is exothermic.

Two different pathways were found for the forward decomposition of ^{P1}LM1. In the first pathway, the system climbs through the transition state ^{P1}TS2, by breaking the O2−O3 bond (2.164 Å) and forming the C1−O3 bond (1.781 Å), with an

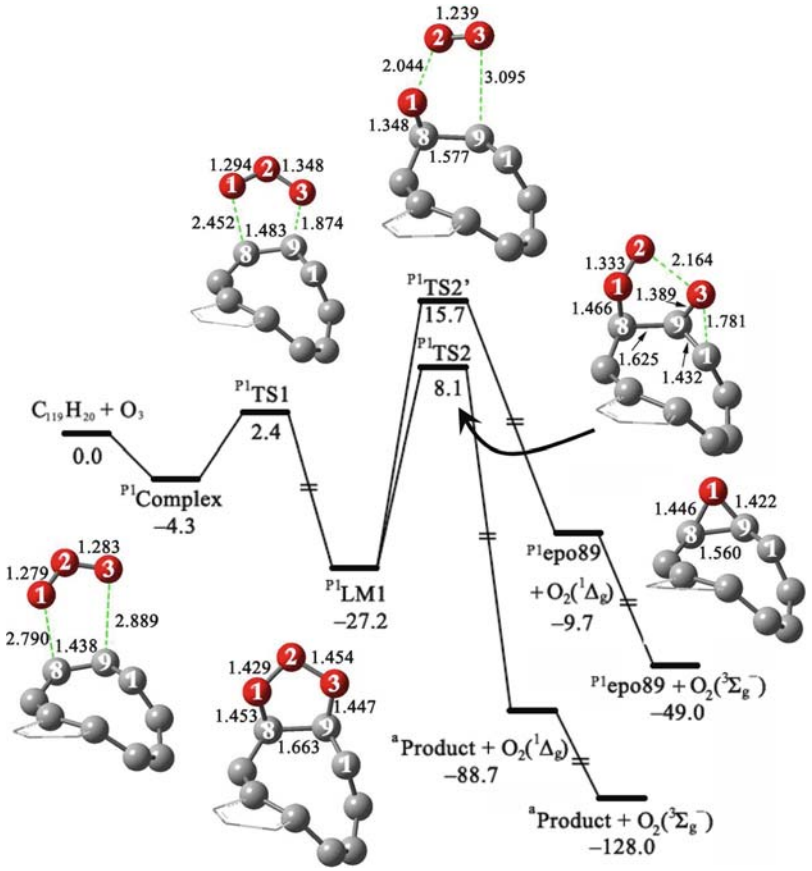


Figure 16-18. Geometries of the reaction center of the transition states, intermediates, and finally product of the reaction of O_3 with the C8−C9 bond (position 1) on the nonagon of $C_{119}H_{20}$. The units of energy and bond length are in kcal/mol and Å, respectively. The oxygen atoms are in red. This figure is adopted from Ref. [184]

activation barrier 35.3 kcal/mol. After overcoming ^{P1}TS2, O1 and O2 dissociate as a singlet O_2, the C9−O3 bond breaks, and O3 forms a carbonylic C=O double bond with C1, and yields the same product ^{a}Product as in the reaction of O_3 on C1. The second decomposition pathway goes through another transition state ^{P1}TS2′ and forms an epoxy adduct ^{P1}epo89. The imaginary vibrational mode clearly shows the breaking of the O1−O2 and C9−O3 bonds. The activation barrier from ^{P1}LM1 to ^{P1}TS2′ is 42.9 kcal/mol, which is 7.6 kcal/mol higher than that of the first decomposition pathway. Contrast to the high exothermicity of the first decomposition pathway (−61.5 kcal/mol relative to ^{P1}LM1), the second decomposition pathway is endothermic (17.5 kcal/mol with respect to ^{P1}LM1). Therefore, the first decomposition pathway is kinetically and thermodynamically much more favourable than the second decomposition pathway. However, the high reaction barrier (35.3 kcal/mol)

from ^{P1}LM1 to the products makes this reaction pathway uncompetitive to the dissociation pathway on the active atom.

16.7.3.3. Reaction of O_3 with the C6−C7 Bond (Position 2)

The structures of ^{P2}Complex, ^{P2}TS1, and ^{P2}LM1 in the course of O_3 attacking the C6−C7 bond are shown in Figure 16-19. ^{P2}Complex is also a π-complex according to its structure. The imaginary frequency mode of ^{P2}TS1 mainly shows the attacking of O1 to C6 and the stretching of the C5−C6 bond. In the primary ozonide ^{P2}LM1, the C5−C6 bond (1.719 Å) is nearly broken. With respect to the energy of the reactants, ^{P2}Complex lays 4.5 kcal/mol lower, whereas ^{P2}TS1 stays 6.4 kcal/mol higher. Thus, the forward reaction barrier from ^{P2}Complex is 10.9 kcal/mol, which

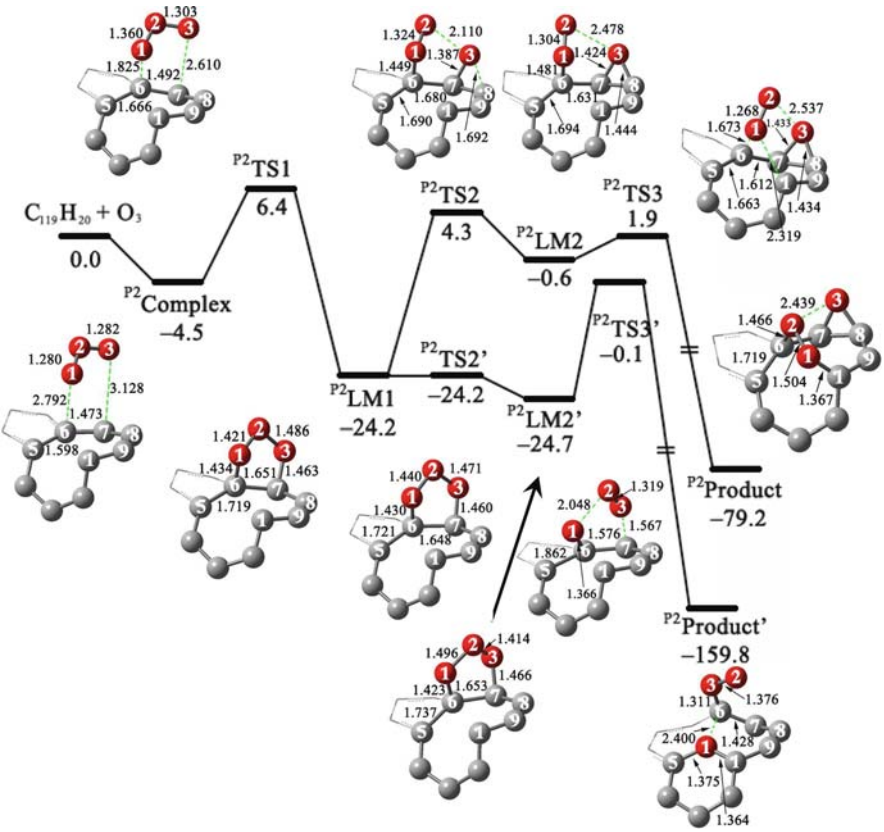


Figure 16-19. Geometries of the reaction center of the transition states, intermediates, and finally product of the reaction of O_3 with the C6−C7 bond (position 2) on the nonagon of $C_{119}H_{20}$. The units of energy and bond length are in kcal/mol and Å, respectively. The oxygen atoms are in red. This figure is adopted from Ref. [184]

is 5.3 kcal/mol higher than that of the reaction on C1. The binding energy of ^{P2}LM1 is −24.2 kcal/mol, thus the reaction of O_3 with C6−C7 is also exothermic.

There are two different decomposition or isomerization pathways of ^{P2}LM1. The first pathway needs to overcome a 28.5 kcal/mol activation barrier and reaches the transition state ^{P2}TS2, in which O3 attacks C8 and breaks the O2−O3 bond. After passing ^{P2}TS2, the system relaxes to ^{P2}LM2 by releasing nearly 5 kcal/mol energy. ^{P2}LM2 is 23.6 kcal/mol less stable than ^{P2}LM1. ^{P2}LM2 can isomerize to ^{P2}Product through another transition state, ^{P2}TS3, by overcoming a small barrier of 2.5 kcal/mol. In ^{P2}TS3, O1 begins to migrate from C6 to C1. In ^{P2}Product, O1 forms bond with C1 and O2 forms bond with C6: O1 and O2 thus coalesce into a bridge between C6 and C1. ^{P2}Product is 79.2 kcal/mol more stable than the reactants.

Another isomerization pathway of ^{P2}LM1 is going through the transition state ^{P2}TS2′, in which O1, O2, O3, C6, and C7 are almost in the same plane. There is virtually no energy cost from ^{P2}LM1 to ^{P2}TS2′. After passing ^{P2}TS2′, the system goes to the endo-primary ozonide ^{P2}LM2′, whose energy is only 0.5 kcal/mol lower than that of the exo-primary ozonide ^{P2}LM1. The isomerization continues moving forward to overcome a 24.8 kcal/mol barrier and reaches ^{P2}TS3′, in which C5−C6 is almost broken (1.862 Å). The imaginary vibrational mode in ^{P2}TS3 shows the breaking of the O3−C7 and O1−O2 bonds and the attacking of O1 to C5. Once overcoming this barrier, the system yields the final product ^{P2}Product′, which is 159.8 kcal/mol more stable than the reactants. Even though, the 24.6 kcal/mol reaction barrier from ^{P2}LM2′ to the final products renders this pathway unfavorable compared to the dissociation on the active atom.

16.7.3.4. Reaction of O_3 with the C4−C5 Bond (Position 3)

^{P3}Complex, in the course of O_3 attacking the C4−C5 bond (Figure 16-20), is also a π-complex according to its structure. The imaginary vibrational mode of transition state right after ^{P3}Complex, ^{P3}TS1, clearly indicates the concerted attacking of O1 to C4 and O3 to C5 and shows the stretching of the C4−C5 bond. The reaction barrier for the reaction on C4−C5 is 19.3 kcal/mol. After overcoming this barrier, the system goes to ^{P3}LM1, which is not a primary ozonide. In ^{P3}LM1, the C4−C5 bond (2.982 Å) is totally broken, and O1, O2, and O3 form a bridge between C4 and C5. ^{P3}LM1 is 46.8 kcal/mol below the energy of the reactants. It is thermodynamically more stable than the primary ozonides ^{P1}LM1 and ^{P2}LM1.

There are two isomerization pathways of ^{P3}LM1. The first one goes through ^{P3}TS2, by breaking of the O1−O2 bond with a barrier only 7.7 kcal/mol. After crossing ^{P3}TS2, the system goes to ^{P3}Product, which is 51.5 kcal/mol more stable than the reactants. Another pathway goes through ^{P3}TS2′, in which O3 attacks C1 and simultaneously breaks the bond with O2 with a barrier of only 8.9 kcal/mol. The product, ^{P3}Product′, in which O3 bridges C1 and C5, is 100.5 kcal/mol more stable than the reactants. In spite of the exothermicity of this reaction pathway, the first reaction barrier from ^{P3}Complex to ^{P3}LM1 is much higher than the corresponding reaction barrier from ^{a}Complex to ^{a}TS1.

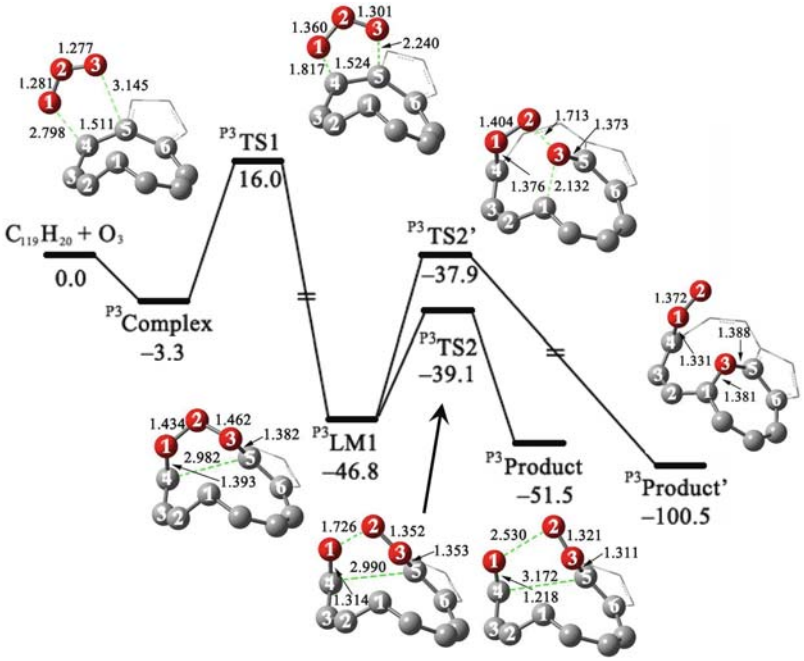


Figure 16-20. Geometries of the reaction center of the transition states, intermediates, and finally product of the reaction of O_3 with the C4−C5 bond (position 3) on the nonagon of $C_{119}H_{20}$. The units of energy and bond length are in kcal/mol and Å, respectively. The oxygen atoms are in red. This figure is adopted from Ref. [184]

16.7.3.5. Reaction of O_3 with the C2−C3 Bond (Position 4)

No π-complex for the reaction on the C2−C3 bond has been found (Figure 16-21). In ^{P4}TS1, the O1−C2 bond is 1.498 Å, which is much shorter than those in the previous initial-reaction transition states. The reaction barrier is 17.3 kcal/mol. The primary ozonide ^{P4}LM1 is only 6.2 kcal/mol more stable than the reactants and is the least stable primary ozonide. A dissociation pathway for ^{P4}LM1 has an activation barrier 23.5 kcal/mol. The system goes through ^{P4}TS2, in which O1 begins to break the bond with O2 and attack C1 at the same time and O2−O3 tends to leave as singlet O_2. After overcoming ^{P4}TS2, O2−O3 indeed dissociate from the system and O1 migrates from C2 to C1, yielding the same final product (^{a}Product) as that in the reaction on C1. The initial reaction barrier is too high for the system to reach ^{P4}LM1 in comparison with the initial reaction barrier of the dissociation at the active carbon atom.

16.7.3.6. Ab initio *Molecular Dynamics Studies*

Above, static quantum mechanical studies have found five different reaction positions of O_3 on the nonagon of the 5-1DB defect site. Among all the reaction pathways, the reaction of O_3 with the active carbon atom C1 is most probable, as it

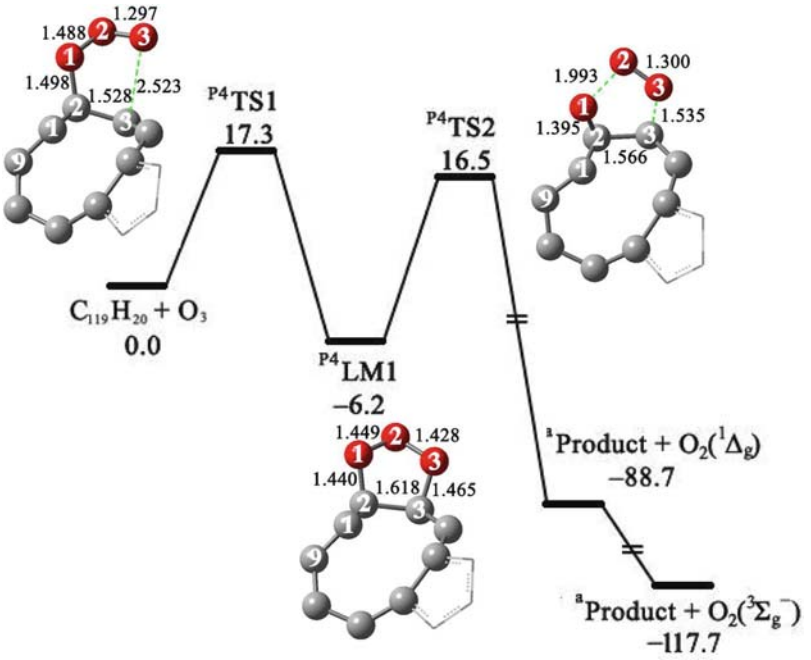


Figure 16-21. Geometries of the reaction center of the transition states, intermediates, and finally product of the reaction of O_3 with the C2−C3 bond (position 4) on the nonagon of $C_{119}H_{20}$. The units of energy and bond length are in kcal/mol and Å, respectively. The oxygen atoms are in *red*. This figure is adopted from Ref. [184]

is a one-step reaction with the lowest initial attacking barrier. This reaction pathway is much more kinetically favorable than other alternatives.

To verify the static quantum mechanical studies of the reactions of O_3 around the 5-1DB defect, ADMP-based AIMD simulations were carried out at 300 K. Initially, an O_3 molecule was placed above the center of the nonagon so that all reactive sites around the 5-1DB defect have equal chance to interact with the incoming O_3 molecule. The dynamical simulation results ratify the spontaneous dissociation of O_3 on C1 to be the most probable reaction process [184]. The change of potential energy of the system during the simulations is shown in Figure 16-22. In less than 50 fs, the system quickly overcomes a *ca.* 20 kcal/mol barrier and releases a large amount of heat (about 160 kcal/mol), during which one oxygen atom is captured by the active carbon atom C1, forming a carbonylic C=O bond, and the other two oxygen atoms leave as O_2. This mechanism is consistent with the scenario from the static quantum mechanical study discussed above, despite the differences in the energies due to the different methods used for the low layer of the 2-layered ONIOM model.

In summary, among the five possible reactive positions of the nonagon of the 5-1DB defect in the vacancy-defected (5,5) SWCNT, the most favorite reaction takes place on the active carbon atom, through a one-step process, in which the

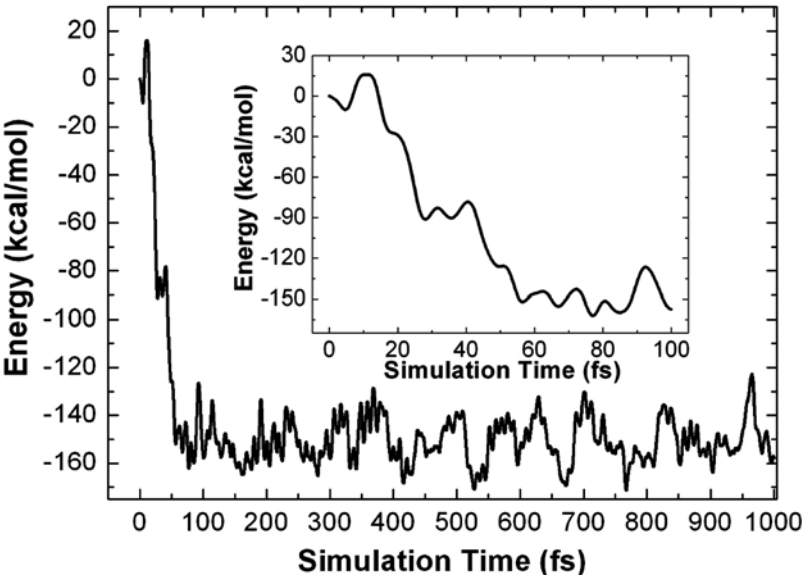


Figure 16-22. The relative potential energy (in kcal/mol) for the system during the ADMP simulation of 1 ps at 300 K. The insert is the relative potential energy for the first 100 fs. This figure is adopted from Ref. [184]

active carbon atom captures an oxygen atom from O_3 and the remaining two oxygen atoms dissociate away as O_2. ADMP dynamical simulation at 300 K confirms the fast spontaneous dissociation of O_3 on the 5-1DB defect.

16.8. CONCLUSIONS AND OUTLOOKS

Electronic properties and reactivities of perfect, defected, and doped SWCNTs have been reviewed based on the theoretical works from us and other groups.

Without defect, a SWCNT with open ends has delocalized electronic structure. Hemispherical caps stabilize SWCNT rods and decrease the HOMO-LUMO gap of SWCNT rod from that of SWCNT segment with open ends. The DOS, LDOS, and FMOs further confirm the existence of localized states on the hemispherical caps in the SWCNT rods. The circular *cis*-polyene chain between the cap and the sidewall of a SWCNT rod is active in chemical reactions according to the MO analysis.

Introduction of vacancy defect brings localized active center to SWCNT (either with open ends or with hemispherical caps) and fullerens. Vacancy defects on SWCNTs also decrease the HOMO-LUMO gaps, destruct the π conjugated system of the FMOs, and enhance their chemical activity. The chemical reactivity of the vacancy defect can be utilized to functionlize SWCNTs and to broaden the applications of SWCNTs in electronic devices, hydrogen storage, chemical sensor, catalysts, and optical materials. Reactions of NO and O_3 with vacancy-defected (5,5) SWCNT showcase such novel functionlizations.

Doping of hetero-atom to SWCNT results in new active center. Depending on the nature of the doping atom, the doped HSWCNT can have very different properties. The adsorption of gas molecules on HSWCNT reveals different chemical activity of the doping site. Such doped HSWCNT can be utilized as catalyst, chemical sensor, and hydrogen storage material because of its chemical sensitivity, reactivity, and selectivity.

Our studies point to new directions for future applications of HSWCNTs in catalysis, chemical sensor, surface science, and nanotube chemistry. Certainly, more sustained research efforts should be carried out to deepen our understanding of the wonders of carbon nanotubes.

ACKNOWLEDGMENTS

The financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada is gratefully acknowledged. L.V.L. received the Gladys Estella Laird Fellowship and the Charles A. McDowell Fellowship from the Department of Chemistry at the University of British Columbia. W.Q.T. thanks the startup fund from Jilin University.

REFERENCES

1. W. Q. Tian, L. V. Liu, and Y. A. Wang, in *Handbook of Theoretical and Computational Nanotechnology*, Vol. 9, edited by M. Rieth and W. Schommers (American Scientific, Valencia, California, USA, 2006), Chap. 10, pp. 499–524.
2. S. Iijima and T. Ichihashi, Nature **363**, 603 (1993).
3. D. S. Bethune, C. H. Kiang, M. S. de Vries, G. Gorman, R. Savoy, J. Vazquez, and R. Beyers, Nature **363**, 605 (1993).
4. S. Niyogi, M. A. Hamon, H. Hu, B. Zhao, P. Bhowmik, R. Sen, M. E. Itkis, and R. C. Haddon, Acc. Chem. Res. **35**, 1105 (2002).
5. T. Hermraj-Benny, S. Banerjee, and S. S. Wong, Chem. Mater. **16**, 1855 (2004).
6. J.-M. Nhut, P. Nguyen, C. Pham-Huu, N. Keller, and M.-J. Ledoux, Catal. Today **91–92**, 91 (2004).
7. J. Zhang, H. Zou, Q. Qing, Y. Yang, Q. Li, Z. Liu, X. Guo, and Z. Du, J. Phys. Chem. B **107**, 3712 (2003).
8. T. Yamabe, M. Imade, M. Tanaka, and T. Sato, Synth. Metals **117**, 61 (2001).
9. X. Lu, F. Tian, Y. Feng, X. Xu, N. Wang, and Q. Zhang, Nano Lett. **2**, 1325 (2002).
10. J. Li, Y. Zhang, and M. Zhang, Chem. Phys. Lett. **364**, 328 (2002).
11. J. Cioslowski, N. Rao, and D. Moncrieff, J. Am. Chem. Soc. **124**, 8485 (2002).
12. T. Kar, B. Akdim, X. Duan, and R. Pachter, Chem. Phys. Lett. **392**, 176 (2004).
13. M. Zhao, Y. Xia, J. P. Lewis, and L. Mei, J. Phys. Chem. B **108**, 9599 (2004).
14. S. Gustavsson, A. Rosén, H. Grennberg, and K. Bolton, Chem. Eur. J. **10**, 2223 (2004).
15. E. Joselevich, Chem. Phys. Chem. **5**, 619 (2004).
16. Z. Zhou, M. Steigerwald, M. Hybertsen, L. Brus, and R. Friesner, J. Am. Chem. Soc. **126**, 3597 (2004).
17. T. Yumura, K. Hirahara, S. Bandow, K. Yoshizawa, and S. Iijima, Chem. Phys. Lett. **386**, 38 (2004).
18. D. L. Carroll, P. Redlich, P. M. Ajayan, J. C. Charlier, X. Blase, A. De Vita, and R. Car, Phys. Rev. Lett. **78**, 2811 (1997).

19. Z. Klusek, P. Kowalczyk, and P. Byszewski, Vacuum **63**, 145 (2001).
20. M. Shiraishi and M. Ata, Synth. Metals **128**, 235 (2002).
21. K. A. Dean and B. R. Chalamala, J. Vac. Sci. Technol. B **21**, 868 (2003).
22. H. Kim, J. Lee, S.-J. Kahng, Y.-W. Son, S. B. Lee, C.-K. Lee, J. Ihm, and Y. Kuk, Phys. Rev. Lett. **90**, 216107 (2003).
23. X. Blase, L. X. Benedict, E. L. Shirley, and S. G.Louie, Phys. Rev. Lett. **72**, 1878 (1994).
24. Y. H. Lee, S. G. Kim, and D. Tománek, Phys. Rev. Lett. **78**, 2393 (1997).
25. T. Yaguchi and T. Ando, J. Phys. Soc. Jpn. **70**, 1327 (2001).
26. T. Yaguchi and T. Ando, J. Phys. Soc. Jpn. **71**, 2224 (2002).
27. J. Jiang, J. Dong, and D. Y. Xing, Phys. Rev. B **65**, 245418 (2002).
28. S. Compernolle, L. Chibotaru, and A. Ceulemans, J. Chem. Phys. **119**, 2854 (2003).
29. L. Chico and W. Jaskólski, Phys. Rev. B **69**, 085406 (2004).
30. G. Y. Guo, K. C. Chu, D.-S. Wang, and C.-G. Duan, Phys. Rev. B **69**, 205416 (2004).
31. P. M. Ajayan and O. Z. Zhou, in *Carbon Nanotubes Synthesis, Structure, Properties, and Applications*, edited by M. S. Dresselhaus, G. Dresselhaus, and P. Avoutis (Springer-Verlag, Berlin, 2001).
32. P. Avouris, Acc. Chem. Res. **35**, 1026 (2002).
33. R. D. Antonov and A. T. Johnson, Phys. Rev. Lett. **83**, 3274 (1999).
34. M. S. Fuhrer, J. Nygård, L. Shih, M. Forero, Y.-G. Yoon, M. S. C. Mazzoni, H. J. Choi, J. Ihm, S. G. Louie, A. Zettl, and P. L. McEuen, Science **288**, 494 (2000).
35. J. Kong, N. R. Franklin, C. Zhou, M. G. Chapline, S. Peng, K. Cho, and H. Dai, Science **287**, 622 (2000).
36. P. G. Collins, K. Bradley, M. Ishigami, and A. Zettl, Science **287**, 1801 (2000).
37. A. Goldoni, R. Larciprete, L. Petaccia, and S. Lizzit, J. Am. Soc. Chem. **125**, 11329 (2003).
38. O. Zhou, H. Shimoda, B. Gao, S. Oh, L. Fleming, and G. Yue, Acc. Chem. Res. **35**, 1045 (2002).
39. W. B. Choi, D. S. Chung, J. H. Kang, H. Y. Kim, Y. W. Jin, I. T. Han, Y. H. Lee, J. E. Jung, N. S. Lee, G. S. Park, and J. M. Kim, Appl. Phys. Lett. **75**, 3129 (1999).
40. P. Serp, M. Corrias, and P. Kalck, Appl. Catal. A **253**, 337 (2003).
41. S. Botti, R. Ciardi, L. De Dominicis, L. S. Asilyan, R. Fantoni, and T. Marolo, Chem. Phys. Lett. **378**, 117 (2003).
42. S. Tatsuura, M. Furuki, Y. Sato, I. Iwasa, M. Tian, and H. Mitsu, Adv. Mater. **15**, 534 (2003).
43. S. Y. Set, H. Yaguchi, Y. Tanaka, and M. Jablonski, J. Lightwave Technol. **22**, 51 (2004).
44. A. G. Rozhin, Y. Sakakibara, M. Tokumoto, H. Kataura, and Y. Achiba, Thin Solid Films **464**, 368 (2004).
45. C. Liu, Y. Y. Fan, M. Liu, H. T. Cong, H. M. Cheng, and M. S. Dresselhaus, Science **286**, 1127 (1999).
46. N. Hamada, S.-I. Sawada, and A. Oshiyama, Phys. Rev. Lett. **68**, 1579 (1992).
47. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, *Science of Fullerenes and Carbon Nanotubes* (Academic, San Diego, 1995), Chap. 19.
48. I. W. Chiang, B. E. Brinson, A. Y. Huang, P. A. Willis, M. J. Bronikowski, J. L. Margrave, R. E. Smalley, and R. H. Hauge, J. Phys. Chem. B **105**, 8297 (2001).
49. E. T. Mickelson, C. B. Huffman, A. G. Rinzler, R. E. Smalley, R. H. Hauge, and J. L. Margrave, Chem. Phys. Lett. **296**, 188 (1998).
50. P. J. Boul, J. Liu, E. T. Mickelson, C. B. Huffman, L. M. Ericson, I. W. Chiang, K. A. Smith, D. T. Colbert, R. H. Hauge, J. L. Margrave, and R. E. Smalley, Chem. Phys. Lett. **310**, 367 (1999).
51. M. Holzinger, O. Vostrowsky, F. H. Hirsch, F. Hennrich, M. Kappes, R. Weiss, and F. Jellen, Angew. Chem. Int. Ed. **40**, 4002 (2001).
52. J. L. Bahr and J. M. Tour, Chem. Mater. **13**, 3823 (2001).

53. J. L. Bahr, J. Yang, D. V. Kosynkin, M. J. Bronikowski, R. E. Smalley, and J. M. Tour, J. Am. Chem. Soc. **123**, 6536 (2001).
54. V. Georgakilas, K. Kordatos, M. Prato, D. M. Guldi, M. Holzinger, and A. Hirsch, J. Am. Chem. Soc. **124**, 760 (2002).
55. P. Umek, J. W. Seo, K. Hernadi, A. Mrzel, P. Pechy, D. D. Mihailovic, and L. Forró, Chem. Mater. **15**, 4751 (2003).
56. J. L. Stevens, A. Y. Huang, H. Peng, L. W. Chiang, V. N. Khabashesku, and J. L. Margrave, Nano Lett. **3**, 331 (2003).
57. H. Peng, L. B. Alemany, J. L. Margrave, and V. N. Khabashesku, J. Am. Chem. Soc. **125**, 15174 (2003).
58. H. Hu, B. Zhao, M. A. Hamon, K. Kamaras, M. E. Itkis, and R. C. Haddon, J. Am. Chem. Soc. **125**, 14893 (2003).
59. S. Banerjee, M. G. C. Kahn, and S. S. Wong, Chem. Eur. J. **9**, 1898 (2003).
60. B. Zhao, H. Hu, and R. C. Haddon, Adv. Funct. Mater. **14**, 71 (2004).
61. L. Zhang, V. U. Kiny, H. Peng, J. Zhu, R. F. M. Lobo, J. L. Margrave, and V. N. Khabashesku, Chem. Mater. **16**, 2055 (2004).
62. M. A. Hamon, J. Chen, H. Hu, Y. Chen, M. E. Itkis, A. M. Rao, P. C. Eklund, and R. C. Haddon, Adv. Mater. **11**, 834 (1999).
63. J. L. Bahr and J. M. Tour, J. Mater. Chem. **12**, 1952 (2002).
64. P. M. Ajayan, V. Ravikumar, and J.-C. Charlier, Phys. Rev. Lett. **81**, 1437 (1998).
65. M. Igami, T. Nakanishi, and T. Ando, J. Phys. Soc. Jpn. **68**, 716 (1999).
66. M. Igami, T. Nakanishi, and T. Ando, Physica B **284**, 1746 (2000).
67. A. V. Krasheninnikov and K. Nordlund, Phys. Solid State **44**, 470 (2002).
68. J.-C. Charlier, Acc. Chem. Res. **35**, 1063 (2002).
69. A. V. Krasheninnikov and K. Nordlund, J. Vac. Sci. Technol. B **20**, 728 (2002).
70. A. J. Lu and B. C. Pan, Phys. Rev. Lett. **92**, 105504 (2004).
71. V. V. Belavin, L. G. Bulusheva, and A. V. Okotrub, Int. J. Quantum Chem. **96**, 239 (2004).
72. L. Valentini, F. Mercuri, I. Armentano, C. Cantalini, S. Picozzi, L. Lozzi, S. Santucci, A. Sgamellotti, and J. M. Kenny, Chem. Phys. Lett. **387**, 356 (2004).
73. D. L. Carroll, P. Redlich, X. Blase, J.-C. Charlier, S. Curran, P. M. Ajayan, S. Roth, and M. Rühle, Phys. Rev. Lett. **81**, 2332 (1998).
74. W. Han, Y. Bando, K. Kurashima, and T. Sato, Chem. Phys. Lett. **299**, 368 (1999).
75. S. Peng and K. Cho, Nano Lett. **3**, 513 (2003).
76. M. Zhao, Y. Xia, J. P. Lewis, and R. Zhang, J. Appl. Phys. **94**, 2398 (2003).
77. A. V. Nikulkina and P. N. D'yachkov, Russ. J. Inorg. Chem. **49**, 430 (2004).
78. P. Lambin, A. A. Lucas, and J. C. Charlier, J. Phys. Chem. Solids **58**, 1833 (1997).
79. H. J. Choi, J. Ihm, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. **84**, 2917 (2000).
80. M. B. Nardelli, J.-L. Fattebert, D. Orlikowski, C. Roland, Q. Zhao, and J. Bernholc, Carbon **38**, 1703 (2000).
81. H.-F. Hu, Y.-B. Li, and H.-B. He, Diam. Relat. Mater. **10**, 1818 (2001).
82. L. G. Zhou and S. Q. Shi, Carbon **41**, 579 (2003).
83. Y. Miyamoto, A. Rubio, S. Berber, M. Yoon, and D. Tománek, Phys. Rev. B **69**, 121413 (2004).
84. D. E. Clemmer, J. M. Hunter, K. B. Shelimov, and M. F. Jarrold, Nature **372**, 248 (1994).
85. W. Branz, I. M. L. Billas, N. Malinowski, F. Tast, M. Heinebrodt, and T. P. Martin, J. Chem. Phys. **109**, 3425 (1998).
86. J. M. Poblet, J. Muñoz, K. Winkler, M. Cancilla, A. Hayashi, C. B. Lebrilla, and A. L. Balch, Chem. Commun. 493 (1999).
87. Q. Kong, Y. Shen, L. Zhao, J. Zhuang, S. Qian, Y. Li, Y. Lin, and R. Cai, J. Chem. Phys. **116**, 128 (2002).

88. A. Hayashi, Y. Xie, J. M. Poblet, J. M. Campanera, C. B. Lebrilla, and A. L. Balch, J. Phys. Chem. A **108**, 2192 (2004).
89. C. Ding, J. Yang, X. Cui, and C. T. Chan, J. Chem. Phys. **111**, 8481 (1999).
90. I. M. L. Billas, C. Massobrio, M. Boero, M. Parrinello, W. Branz, F. Tast, N. Malinowski, M. Heinebrodt, and T. P. Martin, Comput. Mater. Sci. **17**, 191 (2000).
91. X. Blase, J.-C. Charlier, A. De Vita, and R. Car, Appl. Phys. Lett. **70**, 197 (1996).
92. D. Golberg, Y. Bando, W. Han, K. Kurashima, and T. Sato, Chem. Phys. Lett. **308**, 337 (1999).
93. S. L. Sung, S. H. Tsai, C. H. Tseng, F. K. Chiang, X. W. Liu, and H. C. Shih, Appl. Phys. Lett. **74**, 197 (1999).
94. L. V. Liu, W. Q. Tian, and Y. A. Wang, J. Phys. Chem. B **110**, 1999 (2006).
95. A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (Dover, New York, 1996).
96. R. McWeeny, *Methods of Molecular Quantum Mechanics*, 2nd Ed. (Academic, New York, 1992).
97. F. Jensen, *Introduction to Computational Chemistry* (Wiley, New York, 1998).
98. D. C. Young, *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems* (Wiley, New York, 2001).
99. C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models* (Wiley, New York, 2002).
100. P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
101. W. Kohn and L. J. Sham, Phys. Rev. **40**, A1133 (1965).
102. R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).
103. Y. A. Wang and E. A. Carter, in *Theoretical Methods in Condensed Phase Chemistry*, edited by S. D. Schwartz (Kluwer, Dordrecht, 2000), p. 117.
104. C. C. J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).
105. B. O. Roos, in *Advances in Chemical Physics: Ab Initio Mehtods in Quantum Chemistry Part II*, Vol. 69, edited by K. P. Lawley (Wiley, New York, 1987), p. 1.
106. I. Shavitt, in *Modern Theoretical Chemistry*, Vol. 3, edited by H. F. Schaefer III (Plenum, New York, 1977), p. 189.
107. J. Cizek, Chem. Phys. **45**, 4256 (1966).
108. J. A. Pople, D. P. Santry, and G. A. Segal, J. Chem. Phys. **43**, S129 (1965).
109. J. A. Pople and G. A. Segal, J. Chem. Phys. **43**, S136 (1965).
110. J. A. Pople and G. A. Segal, J. Chem. Phys. **44**, 3289 (1966).
111. J. A. Pople, D. L. Beveridge, and P. A. Dobosh, J. Chem. Phys. **47**, 2026 (1967).
112. M. J. S. Dewar and W. Thiel, J. Am. Chem. Soc. **99**, 4899 (1977).
113. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, J. Am. Chem. Soc. **107**, 3902 (1985).
114. J. J. P. Stewart, J. Comput. Chem. **10**, 209 (1989).
115. J. J. P. Stewart, J. Comput. Chem. **10**, 221 (1989).
116. W. Kohn, A. D. Becke, and R. G. Parr, J. Phys. Chem. **100**, 12974 (1996).
117. A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
118. S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, and M. J. Frisch, J. Mol. Struct. (Theochem) **461**, 1 (1999).
119. T. Klüner, N. Govind, Y. A. Wang, and E. A. Carter, J. Chem. Phys. **116**, 42 (2002).
120. T. Klüner, N. Govind, Y. A. Wang, and E. A. Carter, Phys. Rev. Lett. **88**, 209702 (2002).
121. T. Klüner, N. Govind, Y. A. Wang, and E. A. Carter, Phys. Rev. Lett. **86**, 5954 (2001).
122. N. Govind, Y. A. Wang, and E. A. Carter, J. Chem. Phys. **110**, 7677 (1999).
123. N. Govind, Y. A. Wang, A. J. R. da Silva, and E. A. Carter, Chem. Phys. Lett. **295**, 129 (1998).
124. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).

125. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comp. Chem. **4**, 187 (1983).
126. N. L. Allinger, Y. H. Yuh, and J.-H. Lii, J. Am. Chem. Soc. **111**, 8551 (1989).
127. A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff, J. Am. Chem. Soc. **114**, 10024 (1992).
128. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
129. A. D. Mackerell, D. Bashford, M. Bellott, R. I. Dunbrack, J. D. Evanseck, M. J. Field, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, T. F. K. Lau, C. Mattos, S. Michnick, T. Nago, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkievicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
130. A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff, J. Am. Chem. Soc. **114**, 10046 (1992).
131. D. L. Thompson, Ed., *Modern Methods for Multidimensional Dynamics Computation in Chemistry* (World Scientific, Singapore, 1998).
132. R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
133. H. B. Schlegel, J. M. Millam, S. S. Iyengar, G. A. Voth, A. D. Daniels, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **114**, 9758 (2001).
134. S. S. Iyengar, H. B. Schlegel, J. M. Millam, G. A. Voth, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **115**, 10291 (2001).
135. H. B. Schlegel, S. S. Iyengar, X. Li, J. M. Millam, G. A. Voth, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **117**, 8694 (2002).
136. S. S. Iyengar, H. B. Schlegel, G. A. Voth, J. M. Millam, G. E. Scuseria, and M. J. Frisch, Isr. J. Chem. **42**, 191 (2002).
137. S. Bandow, K. Hirahara, T. Hiraoka, G. Chen, P. C. Eklund, and S. Iijima, MRS Bull. **29**, 260 (2004).
138. A. De Vita, J.-C. Charlier, X. Blasé, and R. Car, Appl. Phys. A **68**, 283 (1999).
139. R. Tamura and M. Tsukada, Phys. Rev. B **52**, 6015 (1995).
140. Gaussian 03, Revision B.05, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh, PA, 2003.
141. R. Ditchfield, W. J. Hehre, and J. A. Pople, J. Chem. Phys. **54**, 724 (1971).
142. W. J. Hehre, R. Ditchfield, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972).
143. P. C. Hariharan and J. A. Pople, Mol. Phys. **27**, 209 (1974).
144. M. S. Gordon, Chem. Phys. Lett. **76**, 163 (1980).
145. A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
146. J. P. Perdew, I. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).
147. J. P. Perdew, K. Burke, and Y. Wang, Phys. Rev. B **54**, 16533 (1996).
148. A. E. Reed, L. A. Curtiss, and F. Weinhold, Chem. Rev. **88**, 899 (1988).

149. T. Yumura, S. Bandow, K. Yoshizawa, and S. Iijima, J. Phys. Chem. B **108**, 11426 (2004).
150. J.-P. Deng, D.-D. Ju, G.-R. Her, C.-Y. Mou, C.-J. Chen, Y.-Y. Lin, and C.-C. Han, J. Phys. Chem. **97**, 11575 (1993).
151. M. Saunders, H. A. Jimenezazquez, R. J. Cross, and R. J. Poreda, Science **259**, 1428 (1993).
152. R. L. Murry, D. L. Strout, G. K. Odom, and G. E. Scuseria, Nature **366**, 655 (1993).
153. M.-L. Sun, Z. Slanina, and S.-L. Lee, Fullerene Sci. Technol. **3**, 627 (1995).
154. L. Turker, J. Mol. Struct.: Theochem. **571**, 99 (2001).
155. Y. H. Hu and E. Ruckenstein, J. Chem. Phys. **120**, 7971 (2004).
156. Y. H. Hu and E. Ruckenstein, J. Chem. Phys. **119**, 10073 (2003).
157. S. U. Lee and Y.-K. Han, J. Chem. Phys. **121**, 3941 (2004).
158. J. Ribas-Ariño and J. J. Novoa, Phys. Rev. B **73**, 035405 (2006).
159. A. N. Andriotis, M. Menon, R. M. Sheetz, and L. Chernozatonskii, Phys. Rev. Lett. **90**, 026801 (2003).
160. C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
161. W. Q. Tian, J.-K. Feng, Y. A. Wang, and Y. Aoki, J. Chem. Phys. **125**, 094105 (2006).
162. F. Leclercq, P. Damay, M. Foukani, P. Chieux, M. C. Bellissent-Funel, A. Rassat, and C. Fabre, Phys. Rev. B **48**, 2748 (1993).
163. L. V. Liu, W. Q. Tian, and Y. A. Wang, Int. J. Quantum Chem. **109**, 3441 (2009).
164. S. C. O'Brien, J. R. Heath, R. F. Curl, and R. E. Smalley, J. Chem. Phys. **88**, 220 (1988).
165. C.-H. Kiang, W. A. Goddard III, R. Beyers, and D. S. Bethune, J. Phys. Chem. **100**, 3749 (1996);
166. Y. Zhu, T. Yi, B. Zheng, and L. Cao, Appl. Surf. Sci. **137**, 83 (1999).
167. F. Banhart, Rep. Prog. Phys. **62**, 1181 (1999).
168. J. C. Rienstra-Kiracofe, G. S. Tschumper, and H. F. Schaefer III, Chem. Rev. **102**, 231 (2002).
169. M. Terrones, H. Terrones, F. Banhart, J.-C. Charlier, and P. M. Ajayan, Science **288**, 1226 (2000).
170. D. Srivastava, M. Menon, C. Daraio, S. Jin, B. Sadanadan, and A. M. Rao, Phys. Rev. B **69**, 153414 (2004).
171. A. Hansson, M. Paulsson, and S. Stafstrom, Phys. Rev. B **62**, 7639 (2000).
172. S. B. Fagan, L. B. da Silva, and R. Mota, Nano Lett. **3**, 289 (2003).
173. S. L. Mielke, D. Troya, S. Zhang, J. L. Li, S. P. Xiao, R. Car, R. S. Ruoff, G. C. Schatz, and T. Belytschko, Chem. Phys. Lett. **390**, 413 (2004).
174. M. Sammalkorpi, A. Krasheninnikov, A. Kuronen, K. Nordlund, and K. Kaski, Phys. Rev. B **70**, 245416 (2004).
175. J. Rossato, R. J. Baierle, A. Fazzio, and R. Mota, Nano Lett. **5**, 197 (2005).
176. R. Czerw, M. Terrones, J.-C. Charlier, X. Blase, B. Foley, R. Kamalakaran, N. Grobert, H. Terrones, D. Tekleab, P. M. Ajayan, W. Blau, M. Rühle, and D. L. Carroll, Nano Lett. **1**, 457 (2001).
177. R. H. Xie, Chem. Phys. Lett. **310**, 379 (1993).
178. Y.-H. Kim, Y. Zhao, A. Williamson, M. J. Heben, and S. B. Zhang, Phys. Rev. Lett. **96**, 016102 (2006).
179. E. Hernández, P. Ordejón, I. Boustani, A. Rubio, and J. A. Alonso, J. Chem. Phys. **113**, 3814 (2000).
180. W. Q. Tian, L. V. Liu, and Y. A. Wang, Phys. Chem. Chem. Phys. **8**, 3528 (2006).
181. S. H. Yang, W. H. Shin, J. W. Lee, S. Y. Kim, S. I. Woo, and J. K. Kang, J. Phys. Chem. B **110**, 13941 (2006).
182. P. J. Hay and W. R. Wadt, J. Chem. Phys. **82**, 299 (1985).
183. G. Mpourmpakis, G. E. Froudakis, A. N. Andriotis, and M. Menon, Appl. Phys. Lett. **87**, 193105 (2005).
184. L. V. Liu, W. Q. Tian, and Y. A. Wang, J. Phys. Chem. B **110**, 13037 (2006).
185. A. J. Lu and B. C. Pan, Phys. Rev. B **71**, 165416 (2005).

186. K. Fukui, Science **218**, 747 (1987).
187. W. Q. Tian and Y. A. Wang, J. Chem. Theory Comput. **1**, 353 (2005).
188. N. Rega, S. S. Iyengar, G. A. Voth, H. B. Schlegel, T. Vreven, and M. J. Frisch, J. Phys. Chem. B **108**, 4210 (2004).
189. C. K. Kohlmiller and L. Andrews, J. Am. Chem. Soc. **103**, 2578 (1981).
190. M. L. McKee and C. M. Rohlfing, J. Am. Chem. Soc. **111**, 2497 (1989).

CHAPTER 17

MULTISCALE MODELING OF BIOLOGICAL PROTEIN MATERIALS – DEFORMATION AND FAILURE

SINAN KETEN[1], JEREMIE BERTAUD[1], DIPANJAN SEN[1,2], ZHIPING XU[1], THEODOR ACKBAROW[1], AND MARKUS J. BUEHLER[1]

[1]*Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA, e-mail: mbuehler@MIT.EDU*
[2]*Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA*
Correspondance to M.J. Buehler, mbuehler@MIT.EDU; http://web.mit.edu/mbuehler/www

Abstract: Multi-scale properties of biological protein materials have been the focal point of extensive investigations over the past decades, leading to formation of a research field that connects biology and materials science, referred to as materiomics. In this chapter we review atomistic based modeling approaches applied to study the scale-dependent mechanical behavior of biological protein materials, focused on mechanical deformation and failure properties. Specific examples are provided to illustrate the application of numerical methods that link atomistic to mesoscopic and larger continuum scales. The discussion includes the formulation of atomistic simulation methods, as well as examples that demonstrate their application in case studies focused on size effects of the fracture behavior of protein materials. The link of atomistic scale features of molecular structures to structural scales at length-scales of micrometers will be discussed in the analysis of the mechanics of a simple model of the nuclear lamin network, revealing how protein networks with structural flaws cope with mechanical load

Keywords: Hierarchical material, Nanomechanics, Biological protein materials, Fracture, Deformation, Experiment, Simulation, Materiomics, Multi-scale modeling

17.1. INTRODUCTION

Proteins constitute critical building blocks of life, forming biological protein materials (BPMs) such as hair, bone, skin, spider silk or cells, which play an important role in providing key mechanical functions to biological systems [1–9]. Flaws and failure of these materials can cause serious diseases and malfunctions in biological organisms, for example due to misfolded protein structures. However, the fundamental deformation and failure mechanisms of biological protein materials remain largely unknown, partly due to a lack of understanding of how individual protein

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 473–533.
DOI 10.1007/978-1-4020-9785-0_17,

building blocks respond to mechanical load and how they participate in the function of the overall biological system. Significant advances in experimental, theoretical and computational materials science have enabled a deeper understanding of BPMs through the linking of structure-process-property (SPP). The material properties of BPMs have been the focal point of extensive studies over the past decades, leading to formation of a research field that connects biology and materials science, referred to as materiomics. Materiomics utilizes mechanistic insight, based on BPM SPP relations in its biological context, to provide a basis for understanding disease processes, to develop new approaches to treating genetic and infectious diseases, injury and trauma as well as to enhance engineered materials via translating material concepts from biology.

In biology, structural design and materials engineering is unified through formation of hierarchical features with atomic resolution, from nano to macro (see Figure 17-1, illustrating the multi-scale hierarchical structure of collagenous tissue). Protein materials are capable of unifying disparate properties such as strength (ability to sustain large stresses without fracture) and robustness (ability to undergo deformation without fracture despite the presence of defects, equivalent to the materials science concept of toughness), as well as other dynamical properties such as self-healing ability, adaptability, changeability, and evolvability into

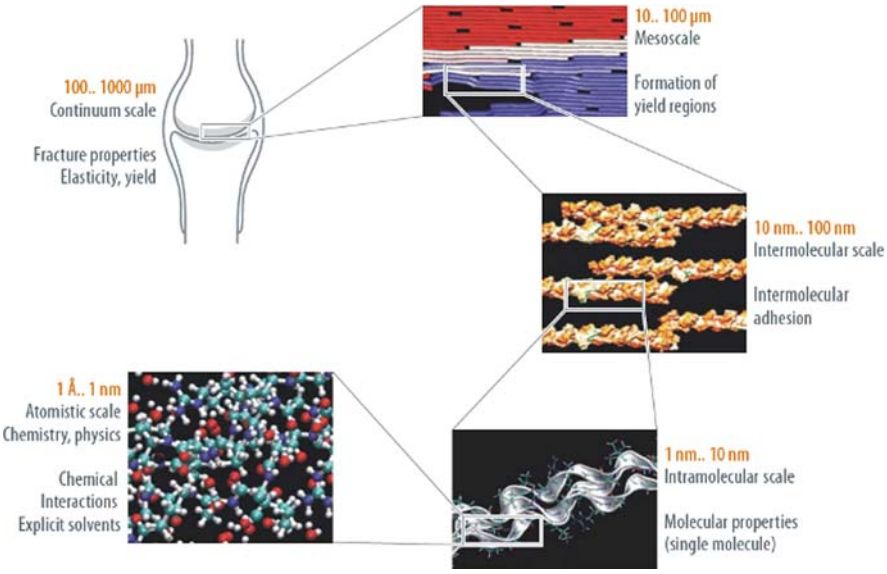


Figure 17-1. Overview over different material scales, from nano to macro, here exemplified for collagenous tissue [15, 17–19]. Biological protein materials such as collagen, skin, bone, spider silk or cytoskeletal networks in cells feature complex, hierarchical structures. The macroscopic mechanical material behavior is controlled by the interplay of properties throughout various scales. In order to understand deformation and fracture mechanisms, it is crucial to elucidate atomistic and molecular mechanisms at each scale (examples are provided in the plot). Computational multi-scale approaches thereby play a crucial role in transcending through multiple scales in length and time

multi-functional materials [1–9]. Many synthetic materials are not capable to unify strength and robustness, leading to materials that are either extremely strong with little ductility and high brittleness (e.g. ceramics, glass, silicon), or weak materials with extreme ductility (e.g. soft metals like copper, nickel or gold). The molecular basis of the ability of protein materials to combine contrasting material properties remains unexplained, albeit it has been suggested that it is perhaps related to the characteristic molecular and hierarchical features found in biological protein materials.

17.1.1. Nanomechanics of Protein Materials: Challenges and Opportunities

The behavior of materials is intimately linked to the atomic microstructure of the material. Whereas crystalline materials show mechanisms such as dislocation spreading or crack extension [10–12], biological materials feature molecular unfolding or sliding, with a particular significance of rupture of chemical bonds such as hydrogen bonds, covalent cross-links or intermolecular entanglement. Additional mechanisms operate at larger length scales, where the interaction of extracellular materials with cells and of cells with one another, different tissue types and the influence of tissue remodeling (at longer time-scales) become more evident. The dominance of specific mechanisms is controlled by geometrical parameters, the chemical nature of the molecular interactions, as well as the structural arrangement of the protein elementary building blocks, across many hierarchical scales, from nano to macro.

A major trait of virtually all biological materials is the occurrence of hierarchies and, at the molecular scale, the abundance of weak interactions. It has been suggested that the presence of hierarchies in biological materials is vital to utilize molecular and sub-molecular features for the generation of biologically functional properties. Material components utilized in the material buildup are sometimes mechanically inferior, such as weak H-bonds or highly brittle mineral crystal phases. Hierarchical arrangements with characteristic length-scales provide the basis to enhance these properties and to generate overall superior mechanical properties. Thereby, hierarchies provide a link between structural organization and function [13], and enable the bridging of nanoscale mechanisms to macroscale properties. The types of processes play a crucial role in understanding the basis of materials in their biological context. For example, the utilization of weak chemical interactions such as H-bonds makes it possible to produce strong materials at moderate temperatures and thus with limited energy use. An important distinction between traditional “engineered” and biological materials is the geometrical occurrence of defects. While defects are often distributed randomly over the volume in crystalline materials, biological materials consist of an ordered arrangement of structure that reaches down to molecular scales. In many biological materials, defects are placed with atomistic or molecular precision, and play a major role in the material behavior observed at larger scales. Examples for such defects include material interfaces between protein and mineral crystals, structural defects in protein domains

(e.g. stutter domains in intermediate filament proteins), or soft protein domains that can undergo large deformation and thereby dissipate great amounts of energy. These features have been observed in bone, nacre, collagenous tissue or cellular protein networks, among others. It is noted that this placement of defects in materials is similar to concepts used in the design of composite materials; however, in biological protein materials structural elements range through many scales, from nano to macro.

The mechanical properties of biological materials have wide ranging implications for biology. In cells for instance, mechanical sensing is used to transmit signals from the environment to the cell nucleus or to control tissue formation and regeneration [1, 14]. The structural integrity and shape of cells is controlled by the cell's cytoskeleton, which resembles an interplay of complex protein structures and signaling cascades arranged in a hierarchical fashion [1]. Bone and collagen, providing structure to our body, or spider silk, used for prey procurement, are examples of materials that have incredible elasticity, strength and robustness unmatched by many synthetic materials, mainly attributed to its structural formation with molecular precision [4, 15–22]. The transfer of concepts observed in biology into technological applications and new materials design remains a big challenge with potential big payoff. In particular, the combination of nanostructural and hierarchical features into materials developments could lead to significant breakthroughs to develop new materials that mimic or exceed the properties found in biological analogs.

The characterization of material properties for biological protein materials may also play a crucial role in developing a better understanding of diseases. Injuries and genetic diseases are often caused by structural changes in protein materials (e.g. defects, flaws, changes to the molecular structure), resulting in failure of the material's intended function. This approach enables one to probe how mutations in structure alter the properties of protein materials. In the case of *osteogenesis imperfecta* (brittle bone disease), for instance, molecular-scale models predict a softening of bone's basic collagen constituent [23]. These observations may eventually provide explanations to the molecular origin of certain diseases. Additionally, these findings provide evidence that material properties play an essential role in biological systems, and that the current paradigm of focusing on biochemistry alone as the cause of diseases is insufficient. It is envisioned that the long-term potential impact of this work can be used to predict diseases in the context of diagnostic tools by measuring material properties rather than focusing on symptomatic chemical readings alone. Such approaches have been explored for cancer and malaria, for instance [24, 25].

17.1.2. Strategy of Investigation

What are the most promising strategies in order to analyze biological protein materials? An integrated approach that uses experiment and predictive simulation concurrently has evolved into a successful research paradigm in materiomics. Experimental techniques have now gained unparalleled accuracy in both length- and time scales (see Figure 17-2), as reflected in development and utilization of

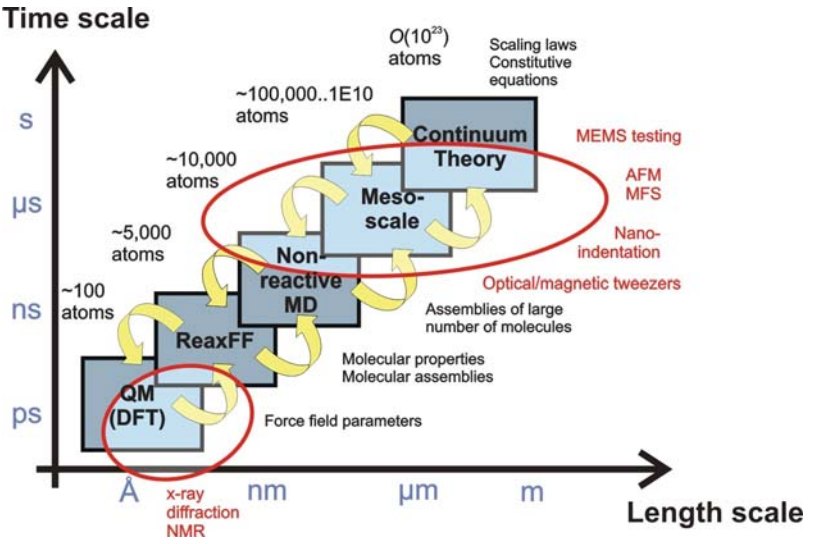


Figure 17-2. Overview over computational and experimental tools. Hierarchical coupling of different computational tools can be used to traverse throughout a wide range of length- and time scales. Such methods enable one to provide a fundamental insight into deformation and fracture phenomena, across various time- and length-scales. Handshaking between different methods enables one to transport information from one scale to another. Eventually, results of atomistic, molecular or mesoscale simulation may feed into constitutive equations or continuum models. While continuum mechanical theories have been very successful for crystalline materials, biological materials require statistical theories. Experimental techniques such as Atomic Force Microscope (AFM), Molecular Force Spectroscopy (MFS), nanoindentation or magnetic/optical tweezers now overlap into atomistic and molecular approaches, enabling direct comparison of experiment and simulation. Techniques such as x-ray diffraction, infrared spectroscopy or NMR provide atomic-scale resolution information about the 3D structure of protein molecules and protein assemblies

Atomic Force Microscope (AFM) [26, 27], magnetic and optical tweezers [28, 29] or nanoindentation [30] to analyze biological protein materials [31]. At the same time, modeling and simulation have evolved into predictive tools that complement experimental analyses (see Figure 17-2) at comparable length- and time-scales.

It is now achievable to start from smallest scales – considering electrons and atoms [32], to reach all the way up to larger mesoscopic and macroscopic scales of biological filaments, fibrils, fibers and entire tissues, by explicitly considering the characteristic structural features at multiple material scales. Even though there are still major challenges ahead of us, this progress is amazing and provides one with a large number of opportunities, transforming materials science as a discipline through increased integration of computational approaches in scientific research.

17.1.3. Impact of Materiomics

The use of a materials science approach to studying biological protein materials may have broader impact beyond the areas identified above. In particular, within the

biological sciences, the field of genomics has advanced our knowledge base through the successful sequencing of entire genomes. In recent years however, extensive efforts have been initiated to move beyond genomics, where fields such as systems biology provide explanation to mechanisms of how genes affect phenotypes and biological function.

The analysis of biological protein materials from a materials science perspective contributes to this realm of efforts by focusing on the material properties of hierarchical multi-scale protein structures. Materiomics investigates the structure-property-link of biological materials, focusing on developing models that quantify stability, loss of stability, and robustness (degree of separation between stability and loss of stability). This general framework can be applied to a variety of material properties (optical, electrical, chemical, and biological). Here we focus on mechanical properties at nano- and mesoscale, as the "mechanics" of chemical bonding provides a general basis for many biologically relevant material properties (Figure 17-3).

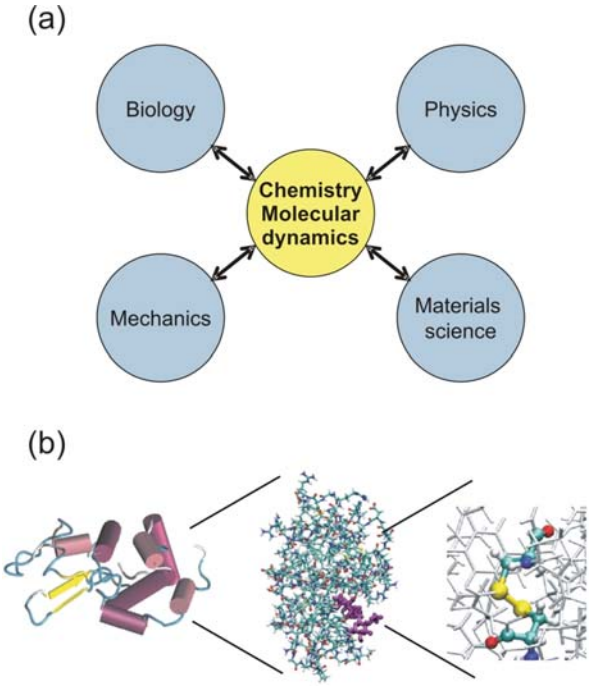


Figure 17-3. Chemistry is the most fundamental "language" of materials science and nanomechanics of biological protein materials. Many other disciplines can link up with the notion of a chemical bond that defines the basic structure and eventually all properties of materials, thereby representing a joint root for these disciplines. Subplot (**a**) shows the central role of chemistry and molecular dynamics schematically. Subplot (**b**) shows for the example of lysozyme how an array of chemical bonds defines the structure of proteins at the most fundamental level (the chemical bonds include covalent bonds (C–C, C–O, C–H, C–N), electrostatic interactions (between charged amino acid side chains), H-bonds (e.g. between H and O or interperpeptide H-bonds in alpha-helices or beta-sheets), as well as vdW interactions (between uncharged parts of molecules)

17.1.4. Transfer from Biological Protein Materials to Synthetic Materials

Biological materials, particularly those used for structural and load-carrying applications, often feature complex hierarchical structures. The constituents of these materials can be soft proteins and hard and brittle minerals, combined in a particular fashion to produce tough strong structures. Studies of their structural morphology and how it contributes to their intriguing properties can provide us with design ideas for using commonly available materials judiciously such as metals or ceramics to obtain strong materials and structures.

A recurring design element in biological hierarchical structures is the prevalence of structural complexity, as seen in several length-scales of organized structure from nano- to macro-scale in a single material, e.g. in bone. The complexity of structure seems to be of great significance in these materials, and has been linked to robustness against perturbations, both environmental and internal, and to the ability of materials to self-heal [22, 33, 34]. This complexity also requires a high degree of self-organization, typically established through self-assembly mechanisms driven by intermolecular interactions of the basic protein constituents.

The importance of hierarchical structures to improving divergent mechanical properties in these materials has also been proposed by several authors. For example, the structural hierarchy seen in skeleton of sea sponge [35–38] has been suggested to be responsible for its high strength and crack resistance, despite being made almost completely of brittle silica constituents. Hierarchical arrangements in protein structures from amino acids up to secondary structures have been proposed as an arrangement for improving robustness of protein materials [39–41].

The ability to transfer hierarchical design from biological to synthetic materials may eventually allow us to extend these properties to conventional manufacturing materials [42]. Through this approach, metals or ceramics can be developed that possess contrasting properties such as high stiffness and toughness, and that can simultaneously be robust against damage caused by large external mechanical loads.

17.2. ATOMISTIC SIMULATION METHODS

In the following sections we briefly review basic atomistic and molecular simulation approaches, focusing on molecular dynamics simulation, a selection of force fields, and a brief discussion of multi-scale approaches through coarse-graining.

17.2.1. Molecular Dynamics Formulation

Atomistic molecular dynamics (MD) is a suitable tool for elucidating the atomistic mechanisms that control deformation and rupture of chemical bonds at nano-scale, and to relate this information to macroscopic materials failure phenomena (see, e.g. review articles and books [43–45], and recent articles from our group that describes large-scale MD simulation of brittle fracture mechanisms [46–50]). The basic concept behind atomistic simulation via MD is to calculate the dynamical trajectory of each atom in the material, by considering their atomic interaction potentials,

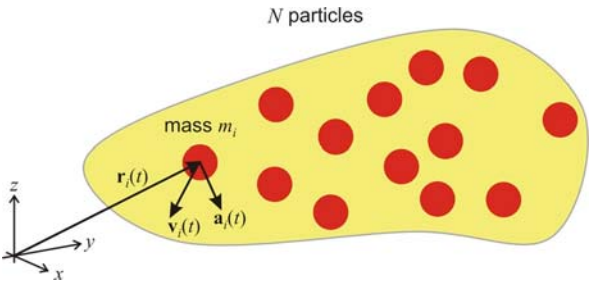


Figure 17-4. Basic formulation of molecular dynamics. This numerical approach predicts the position, velocity and force (acceleration) vector of a set of *N* particles in a system

by solving each atom's equation of motion according to $F = ma$, leading to positions $r_i(t)$, velocities $v_i(t)$ and accelerations $a_i(t)$. The basic approach is shown in Figure 17-4. The numerical integration of Newton's law by considering proper interatomic potentials to obtain interatomic forces enables one to simulate a large ensemble of atoms that represents a larger material volume, albeit typically limited to several nanoseconds of time scale. The availability of interatomic potentials for a specific material (based on the characteristic type of chemical bonding) is often a limiting factor for the applicability of this method.

Classical molecular dynamics generates the trajectories of a large number of particles, interacting with a specific interatomic potential. Thereby, the complex 3D structure of an atom (composed of electrons and a core of neutrons and protons) is approximated by a point particle (Figure 17-5a). Molecular dynamics is an alternative approach to methods like Monte-Carlo, with the distinction that MD provides full dynamical information and deterministic trajectories, which is crucial to describe highly driven phenomena such as catastrophic failure of a protein domain due to laterally applied loads. It is emphasized that Monte-Carlo schemes provide certain advantages as well; however, this point will not be discussed further here as most simulation studies reviewed here are carried out with a MD approach. The total energy of the system is written as the sum of kinetic energy (K) and potential energy (U),

$$E = K + U, \tag{17-1}$$

where the kinetic energy is

$$K = \frac{1}{2} m \sum_{j=1}^{N} v_j^2, \tag{17-2}$$

and the potential energy is a function of the atomic coordinates r_j,

$$U = U(r_j), \tag{17-3}$$

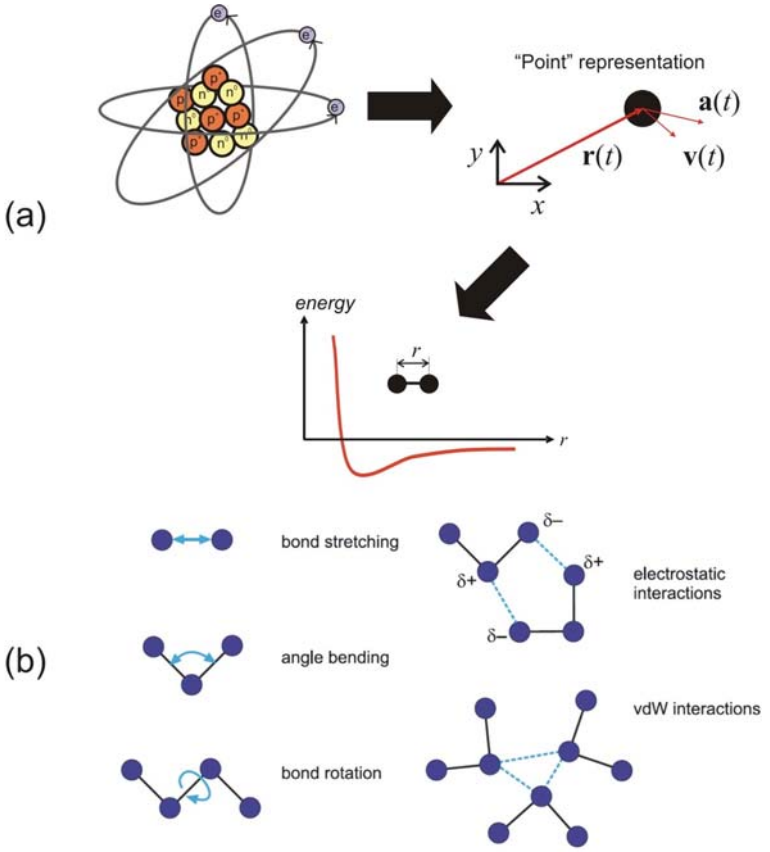


Figure 17-5. Illustration of the concept of point representation as used in molecular dynamics and development of force field models for organic molecules. Subplot (**a**): Effective interatomic potentials (as shown in lower part of the figure) provide a description of the energy landscape of the atoms, capturing the overall effect of the interactions of the atoms due to their specific quantum mechanical structure (electrons, neutrons, protons, as shown in the upper left part). Subplot (**b**): Illustration of how the chemical bonding characteristics is modeled by decomposing into energy contributions due to bending, stretching, rotation and other interactions (similar as the approach used in the definition of the CHARMM model and other related force fields)

with a properly defined potential energy surface $U(r_j)$. The numerical problem to be solved is a system of coupled second order nonlinear differential equations:

$$m\frac{d^2 r_j}{dt^2} = -\nabla_{r_j} U(r_j) \quad j = 1..N, \tag{17-4}$$

which can only be solved numerically for more than two particles, $N > 2$. Typically, MD is based on updating schemes that yield new positions from the old positions, velocities and the current accelerations of particles:

$$r_i(t_0 + \Delta t) = -r_i(t_0 - \Delta t) + 2r_i(t_0)\Delta t + a_i(t_0)\,(\Delta t)^2 + \ldots . \tag{17-5}$$

The forces and accelerations are related by $a_i = f_i/m$. The forces are obtained from the potential energy surface – sometimes also called force field – as

$$F = m\frac{d^2 r_j}{dt^2} = -\nabla_{r_j} U(r_j) \quad j = 1...N. \tag{17-6}$$

This technique can also be used for not only single atoms but also groups of atoms as in the case of coarse-grained meso-scale approaches. Provided interatomic potentials are available, MD is capable of directly simulating a variety of materials phenomena, for instance the response of an atomic crystal lattice to applied loading under the presence of a crack-like defect, or the deformation mechanisms of biological molecules including nucleic acids and proteins.

One of the strengths and a unique feature of atomistic methods is its very fundamental viewpoint of materials phenomena. The only physical law that is put into the simulations is Newton's law and a definition of how atoms interact with each other. Despite this very simple basis, very complex phenomena can be simulated. Unlike many continuum mechanics approaches, atomistic techniques require no a priori assumption on the defect dynamics. Once the atomic interactions are chosen, the complete material behavior is determined. Choosing appropriate models for interatomic interactions provides a rather challenging and crucial step that remains subject of a very active discussion in the scientific community. A variety of different interatomic potentials are used in the studies of biological materials at different scales, and different types of protein structures may require the use of different atomistic models. A drawback of atomistic simulations is the difficulty of analyzing results and the large computational resources necessary to perform the simulations. Due to computational limitations, MD simulations are restricted with respect to the time scales that can be reached, limiting overall time spans in such studies to tens of nanoseconds, or in very long simulation studies to fractions of microseconds. Therefore, many MD simulation results of dynamically stretching protein molecules, for instance, have been carried out at large deformation rates, exceeding several m/s.

Recent advances in computational power now enable the simulation of billions of particles in MD simulations, reaching dimensions on the order of micrometers. Figure 17-6 depicts the historical development of computational power over the past decades.

In the next two sections, we provide a brief review of popular interatomic force fields and modeling approaches suitable for simulating the behavior of protein structures. We refer the reader to more extensive review articles for additional information, in particular regarding force field models [51–54].

17.2.2. CHARMM Force Field

All-atom force fields are predominantly used in molecular dynamics simulations of biological materials at the nanoscale as they generally are the most reliable yet computationally efficient way of studying dynamics of macromolecules. A wide

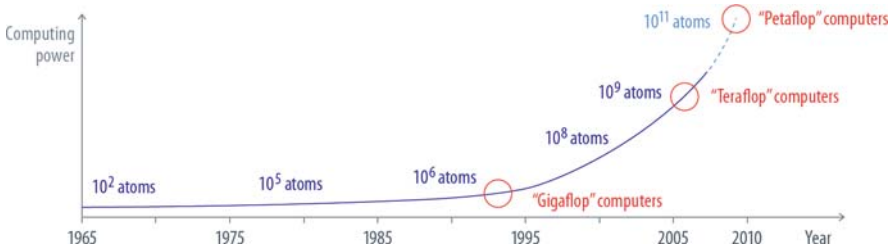


Figure 17-6. Development of computing power over the past decades. The development illustrates the emergence of petaflop computers in the next few years. The plot also summarizes the number of atoms that can be treated with these computing systems; these numbers are developed for simple interatomic potentials with short cutoffs. For CHARMM, the number of atoms is significantly smaller figure. Adopted from Ref. [45]

range of force fields and simulation programs are currently available, most notably the AMBER, the CHARMM force fields and programs, the OPLS force field, the GROMOS/GROMACS [55] packages are commonly used in all-atom molecular dynamics. The NAMD [56] program is a popular code that can carry out computations using CHARMM and other force fields. For the sake of brevity, the main aspects of the CHARMM force field and its implementation in NAMD will be discussed here; the basic concepts of the MD technique and force field formulations are common to all packages used in the field (for a general review, see for instance [52, 57].

The CHARMM force field is widely used in the protein and biophysics community, and provides a reasonable description of the behavior of proteins. The parameters in force fields are often determined from more accurate, quantum chemical simulation models by using the concept of force field training [32] (see the hierarchical coupling schematic shown in Figure 17-2). Parameters for the CHARMM force field have been meticulously optimized and revised over the years taking into consideration a wide variety of input including ab initio results, experimental crystal structures and geometries, as well as vibrational spectra [58].

The potential includes bonding and non-bonding (interaction) terms to describe short and long-range forces between particles. In the CHARMM model, the mathematical formulation for the empirical energy function that contains terms for both internal and external interactions has the form:

$$\begin{aligned} U(\vec{R}) = & \sum_{\textbf{bonds}} K_b(b-b_0)^2 + \sum_{\textbf{UB}} K_{UB}(S-S_0)^2 + \sum_{\textbf{angle}} K_\theta(\theta-\theta_0)^2 + \\ & \sum_{\textbf{dihedrals}} \mathbf{K}_\chi(1+\cos(n\chi-\delta)) + \sum_{\textbf{impropers}} \mathbf{K_{imp}}(\phi-\phi_0)^2 + \\ & \sum_{\textbf{nonbond}} \varepsilon\left[\left(\frac{R_{\min(i,j)}}{r_{ij}}\right)^{12} - \left(\frac{R_{\min(i,j)}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\varepsilon_1 r_{ij}} \end{aligned} \tag{17-7}$$

where K_b, K_{UB}, K_θ, K_χ, and K_{imp} are the bond, Urey-Bradley, angle, dihedral angle, and improper dihedral angle force constants, respectively; b, S, θ, χ and ϕ are the

bond length, Urey-Bradley 1,3-distance, bond angle, dihedral angle, and improper torsion angle, respectively, with the subscript zero representing the equilibrium positions for the individual terms. Figure 17-5b shows a schematic of the individual energy contributions listed in Eq. (17-7).

The Coulomb and Lennard-Jones 6-12 terms constitute the external or nonbonded interactions; ε is the Lennard-Jones well depth and $R_{\min(i,j)}$ is the distance at the Lennard-Jones minimum, q_i is the partial atomic charge, ε_1 is the effective dielectric constant, and r_{ij} is the distance between atoms i and j. In the CHARMM force field, no additional terms are used for H-bonds, since the combination of charge and Lennard-Jones contributions were verified to be adequate for describing protein, solvent and interface hydrogen bonding. In all-atom force fields, water molecules are generally also treated explicitly. Parameters of the force field generally are specified considering a specific water model (e.g. TIP3P dimer model for CHARMM) [52, 57].

The CHARMM force field belongs to a class of models with similar descriptions of the interatomic forces; other models include the DREIDING force field [59], the UFF force field (="Universal Force Field") [60], or the AMBER model [51, 61]. In CHARMM and other classical force fields, bonded terms are modeled with harmonic springs or its variations, and therefore cannot be modified (e.g. towards a different chemical state, such as from sp^2 to sp^3) or broken once defined by the connectivity input obtain from the topology of the molecule. Further, the atomic charges are fixed and cannot change during a simulation. These simplifications improve the simulation speed drastically and are not a major issue for most simulations studying conformational changes of proteins under ambient physiological conditions. On the other hand, simulations in extreme conditions such as mechanical perturbations (e.g. protein unfolding studies) or harsh chemical environments require reactive force fields that can take into account changes in fixed charges of the molecules, formation/breaking of new bonds and variations in bond order.

17.2.3. ReaxFF Force Field

Reactive force fields represent a milestone in overcoming the limitations of classical force fields: Their lack of the ability to describe rupture and formation of covalent bonds. This is because the covalent bond terms are described using harmonic terms, which do not provide an accurate description of the bond energetics at large bond stretch. For mechanical properties of materials (that is, the large-deformation and rupture mechanisms), this translates into the properties of molecules at large-strain, a phenomenon referred to as hyperelasticity (these effects can have profound impact for materials failure mechanisms, as illustrated in [47, 62] for crystalline materials). Figure 17-7a and b illustrates this effect, explaining how nonreactive force fields are not capable of describing transition state energies during bond formation and rupture.

Several flavors of reactive potentials have been proposed in recent years [63–65]. Reactive potentials can overcome the limitations of empirical force fields and enable large-scale simulations of thousands of atoms with quantum mechanics accuracy.

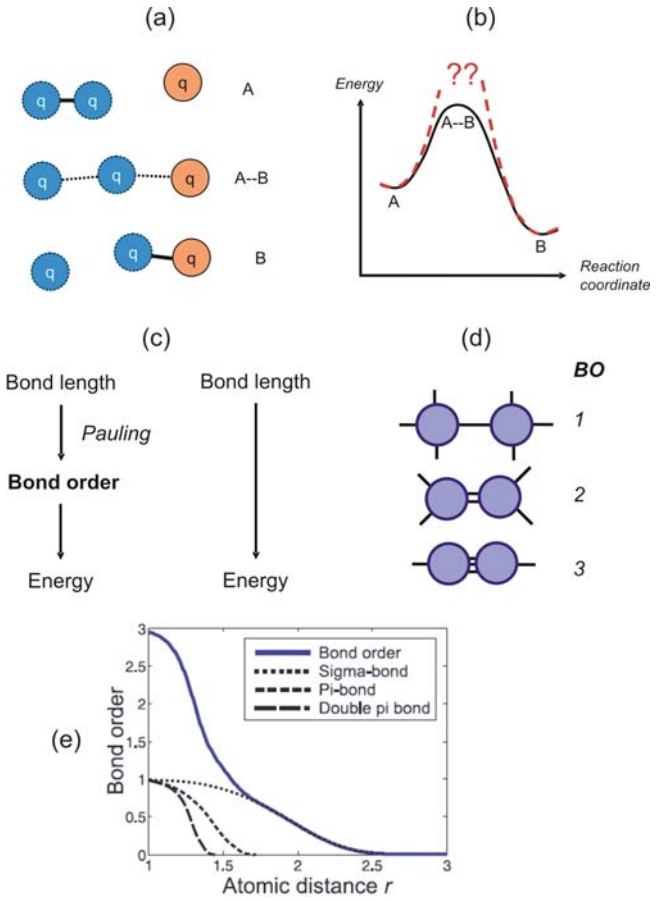


Figure 17-7. Basic concept of the reactive force field formulation (figures adapted from references [63] and [74]). Subplots (**a**) and (**b**) displays the difference between reactive and nonreactive descriptions, showing the ability of reactive models to describe the energy of transition states. Subplot (**c**) shows the concept of defining all energies based on bond orders rather than developing direct relationships between the bond length and energy. Subplot (**d**) illustrates the concept of bond orders, here in an example for a C–C bond. Depending on the distance between atoms, different bond orders are obtained, through a bond order mapping function. This enables to distinguish different quantum chemical states such as sp^3 (single bond), sp^2 (double bond) and sp (triple bond). The continuous change of bond orders as a function of distance ensures that reactive force fields are energy continuous, which is critical to carry our constant energy simulations. At large distances, the bond order vanishes, indicating breaking of the covalent bond. In ReaxFF, the spring constant that characterizes the strength of atomic bonding is modulated by the bond order, leading to vanishing bond strength or dissociation at large stretch. The continuous mapping function is displayed in subplot (**e**)

The reactive potentials, originally only developed for hydrocarbons [49, 50, 66–74], have been extended recently to cover a wide range of materials, including metals, semiconductors and organic chemistry in biological systems such as proteins [63]. Here we focus in particular on the ReaxFF formulation [63, 71]. In some of the

studies reviewed in this article, a particular flavor of the ReaxFF potentials as suggested in [75], with slight modifications to include additional QM data suitable for protein modeling is used.

Reactive potentials are based on a more sophisticated formulation than most nonreactive potentials. A bond length-bond order relationship is used to obtain smooth transition from non-bonded to single, double, and triple bonded systems, as shown in Figure 17-7c–e. All connectivity-dependent interactions (that means, valence and torsion angles) are formulated to be bond-order dependent. This ensures that their energy contributions disappear upon bond dissociation so that no energy discontinuities appear during reactions. The reactive potential also features non-bonded interactions (shielded van der Waals and shielded Coulomb). The reactive formulation uses a geometry-dependent charge calculation (QEq) scheme [76] that accounts for polarization effects and modeling of charge flow, assigning a partial charge to each atom at each integration step. This is considered a critical advance leading to a new bridge between QM and empirical force fields. All interactions feature a finite cutoff distance for computational efficiency, and all interactions are tapered off smoothly at the cutoff distance.

Further discussion regarding reactive force fields for proteins can be found elsewhere [74], including a selection of examples that illustrate the differences between nonreactive and reactive models in describing protein unfolding. For instance, conventional nonreactive models are not able to describe rupture of disulfide cross-links between different protein domains (exemplified for a lysozyme protein structure). In contrast, the reactive model is capable of describing these processes.

The inclusion of covalent bond breaking into the model has major implications on the resulting force-extension curves, as shown in Figure 17-8. This figure shows the force-extension profile during unfolding of a small protein α-conotoxin PnIB from *conus pennaceus* (PDB ID 1AKG), comparing a nonreactive CHARMM model (blue curve) with the reactive ReaxFF model (red curve) [74]. Clearly, the CHARMM model can not capture the bond breaking events at large deformation, and deviates significantly from the ReaxFF description. Similar studies have been carried out for unfolding of lysozyme, as reported in reference [74].

17.2.4. Coarse-Graining Approaches of Protein Structures

Albeit being a very accurate description of macromolecules, all-atom modeling approaches have historically been prohibitively extensive when large systems and long simulation times must be considered. This lead to the development of coarse-grained models [77] (see Figure 17-9), which provide simplified representations of macromolecules employing less degrees of freedom and simple bonded and non-bonded interactions that can be rapidly calculated in each time step. Coarse-grained models have so far been successfully applied to a wide range of problems including protein folding, allostery, aggregation, molecular biomechanics as well as multi-scale description of complex materials such as bone. The various approaches are briefly reviewed here.

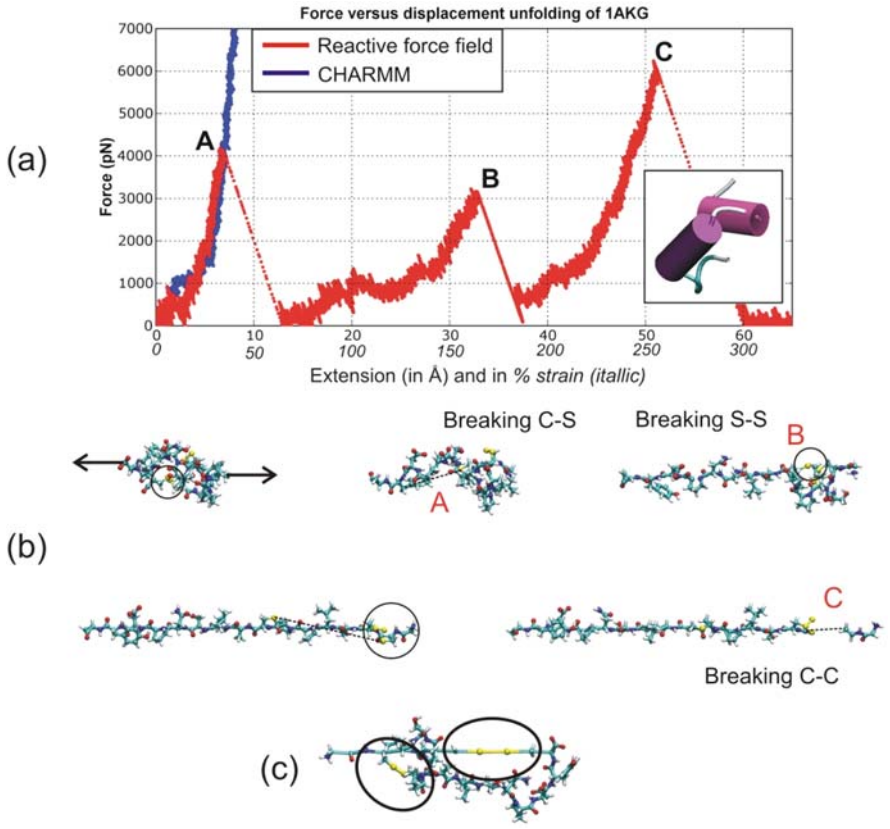


Figure 17-8. Force-extension profile (unfolding of a small protein α-conotoxin PnIB from *conus pennaceus* (PDB ID 1AKG)), comparing a nonreactive CHARMM model (*blue curve*) with the reactive ReaxFF model (*red curve*) [74]. It is apparent that although the ReaxFF and CHARMM descriptions agree for small deformation (below ≈ 7 Å), they disagree strongly for larger deformation. The difference can be explained based on the fact that the CHARMM potential is incapable of describing breaking of the disulfide bonds

17.2.4.1. Single-Bead Models

Single bead models are perhaps the earliest approach taken for studying macromolecules. The term single-bead derives from the idea of using single beads (masses) for describing each amino acid in a protein structure. Elastic Network Model (ENM) [78], Gaussian Network Model (GNM) [79] and Go-model [80] are well known examples of this simplistic approach.

Simple models such as ENM and Go-like models treat each amino acid as a single bead located at the C_α position with mass equal to the mass of the amino acid. The beads are interconnected by harmonic or nonlinear springs representing the covalently bonded protein backbone. In the Go-like models, an additional Lennard-Jones term is included in the potential to describe short-range non-bonded

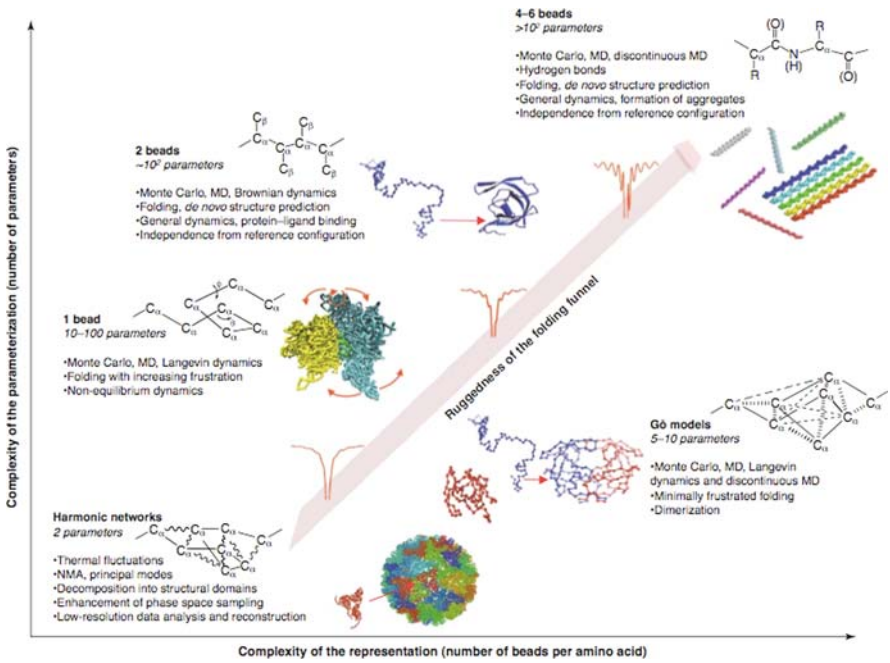


Figure 17-9. Overview over various coarse-graining techniques [77]. The figure illustrates commonly used coarse-graining applications in biomolecular simulations. A schematic representation of the model, indicative number of parameters, methods of solution, main characteristics and applications are shown. Axes indicate increasing complexity of the models in parameterization and molecular representations. Figure reprinted from "Current Opinions in Structural Biology", "Coarse-grained models for proteins", Vol. 15, pp. 144–150, Copyright © 2005, with permission from Elsevier [77]

native interactions between atoms within a cutoff distance. Despite their simplicity, these models have been extremely successful in explaining thermal fluctuations of proteins [77] and have also been implemented to model the unfolding problem to elucidate atomic-level details of deformation and rupture that complement experimental results [81–83]. A more recent direction is coupling of ENM models with a finite element-type framework for mechanistic studies of protein structures and assemblies [84].

Due their simplicity, single-bead models have several shortcomings. With classic ENM, only harmonic deviations from the initial configuration are possible. In the Go-model, native interaction definitions lead to a minimally frustrated landscape which his highly biased towards the input configuration of the molecule. Such models can therefore not predict folding/unfolding intermediates and meta-stable states. The explicit treatment of protein-solvent interactions, nonnative interactions and H-bonds is also not possible with single-bead models. It is now widely accepted that for protein unfolding studies, the results obtained using such models are only qualitative at best, although they may reveal important aspects of topology dependent

mechanical resistance [81–83], and can thus be used to improve our understanding of structure-property links.

17.2.4.2. Multi-Bead Models

Using more than one bead per amino acid can lead to a more detailed description of macromolecules. In the simplest case, the addition of another bead can be used to describe specific side-chain interactions [85]. Four to six bead models capture even higher amount of detail by explicit or united atom description for backbone carbon atoms, side chains, carboxyl and amino groups of amino acids. A great example of this approach is the coarse-grained models developed in Carol Hall's group for studying folding and aggregation in proteins using discontinuous molecular dynamics [86, 87].

Although multi-bead models have superior qualities compared to single-bead descriptions, dozens of additional energetic terms involving pseudobonds and other means to avoid complex dihedral/improper potentials that stabilize the conformation of the polypeptide chain have to be introduced for generic models (see Figure 17-9). Even with the introduction of these terms, some of which are physically not intuitive, the models offer limited applicability, as specific side chain interactions are only valid for simple residues such as glycine and alanine. More complex yet computationally efficient potentials that intrinsically take into account sequence specificity are extremely challenging to develop, thus making readily available all-atom descriptions and simulation packages more favorable for most applications.

17.2.4.3. Coarser Models

More recently, coarser-level modeling approaches have been applied to model biomolecular systems at larger time- and length-scales. These models typically employ superatom descriptions that treat clusters of amino acids as "beads", as shown schematically in Figure 17-10 (for the case of tropocollagen molecules). In such models, the elasticity of the polypeptide chain is captured by simple harmonic or anharmonic (nonlinear) bond and angle terms. These methods are computationally quite efficient and capture shape dependent mechanical phenomena in large biomolecular structures [88], and can also be applied to collagen fibrils in connective tissue [89] as well as mineralized composites such as nascent bone [90].

17.2.4.4. Implicit Solvent

Most biomolecular simulations have either an implicit or explicit treatment of solvent around the molecule, since performing simulations in vacuum can provide misleading results since viscous and hydrophobic effects and dielectric screening properties of solvent are not appropriately described. In particular, the explicit treatment of water is a severe obstacle against scaling up in biological simulations, since a large water box or water sphere needs to be used to keep the protein away from

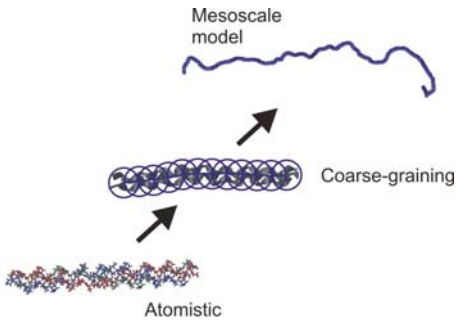


Figure 17-10. Illustration of coarse-graining approach for a simple one-dimensional fibrillar protein filament. This schematic illustrates how a full atomistic representation is coarse-grained and used in a mesoscale model formulation. As shown in Figure 17-3, the mesoscale model formulation enables one to reach much larger time- and length-scales. The systematic parameterization from the bottom up provides a rigorous link between the chemical structure of proteins (for example, through their amino acid sequence) and the overall functional material properties. This computational approach is a key component in the advancement of materiomics as it provides us with the ability to reach microsecond- and micrometer length-scales

fluctuating boundaries. This means that most of the computational effort goes into simulating the motion of the solvent, albeit typically the behavior of the solute (e.g. the protein) is of more interest. This has lead to the development of implicit solvent models, which treat water as a continuum and calculate its effective influence on the solute. In certain cases, implicit solvent techniques can be as much as 50% slower than simulations performed in vacuum. However, simulations done in vacuum are generally hundreds of times faster than explicit solvent runs, thus providing an overall highly efficient approach with significant speedups. These methods take into account free energy of solvation by calculating the accessible surface area of the protein. Such methods can be used in combination with the generalized Born (GB) formulation (numerical approximation to the exact Poisson-Boltzmann equation that is solved in a continuum electrostatic model) to accurately capture the electrostatic and non-polar effects of water on folding mechanisms of proteins [91]. The viscosity of water molecules can also be implemented in such implicit solvent simulations by implementing Langevin dynamics. Recent developments in the analytical treatment of solvent have rendered implicit solvation a reasonable alternative to explicit treatment of water. For more information on the mathematical basis and further details of different solvent models, we refer the reader to the more comprehensive literature reviews in this field [57, 92, 93].

17.2.4.5. Case Study: Coarse-Grained Model of Alpha-Helical Protein Domains

Alpha-helical (AH) protein domains are the key building blocks of cytoskeletal networks as well as hair, hoof and wool. Here we review the development of a coarse-grained model for alpha-helical protein domains with parameters derived

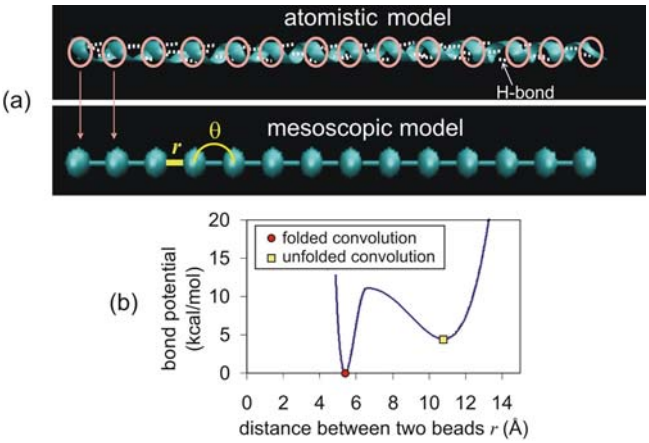


Figure 17-11. Coarse-graining model development of an alpha-helix [94]. Subplot (**a**) shows a schematic of the coarse-graining procedure, changing the full atomistic representation arriving at the mesoscopic bead model. One bead represents one turn (also called convolution) of the alpha-helix so 3.6 residues and has the same mass. Subplot (**b**) shows the two well profile of the bond potential for the mesoscale model. The equilibrium states and energy barriers are obtained from full atomistic simulations. The transition state (peak of the potential) corresponds to the breaking of the 3.6 H-bonds between two convolutions of the alpha-helix. In pulling experiments, after failure of these weak bonds, the convolution unfolds to a second equilibrium state with larger interbead distance, and under further loading, its covalent bonds begin to be stretched which leads to a second increase of the potential

from full atomistic simulations [94]. To achieve the coarse-grained description, the entire sequence of amino acids that makes up the alpha helical structure is replaced by a collection of beads (see Figure 17-11a). In this model, one bead represents one turn (also called convolution) of the alpha-helix, so 3.6 residues, and has the mass corresponding to the one of the protein segment it represents. The beads interact according to a bond potential and an angle potential. The parameters of these interactions are determined from full atomistic simulations of tensile and bending loadings of an alpha helix protein domain.

Figure 17-11b shows the interbead bond potential profile used for the mesoscale model. The potential is a two well potential. The equilibrium states and energy barriers are obtained from full atomistic simulations of tensile loadings. The first equilibrium (first potential minimum) corresponds to the folded state of the turn under no force. The transition state (peak of the potential between two wells) corresponds to the breaking of the 3.6 H-bonds between two turns of the alpha-helix. In pulling experiments, after failure of these weak bonds, the turn unfolds to a second equilibrium state. This corresponds to the second potential minimum with larger interbead distance. Under further loading, its backbone bonds begin to be stretched which leads to a second increase of the potential. Unlike the bond potential, the angle potential profile is a simple harmonic potential. The angle potential constant is determined from bending loading simulations.

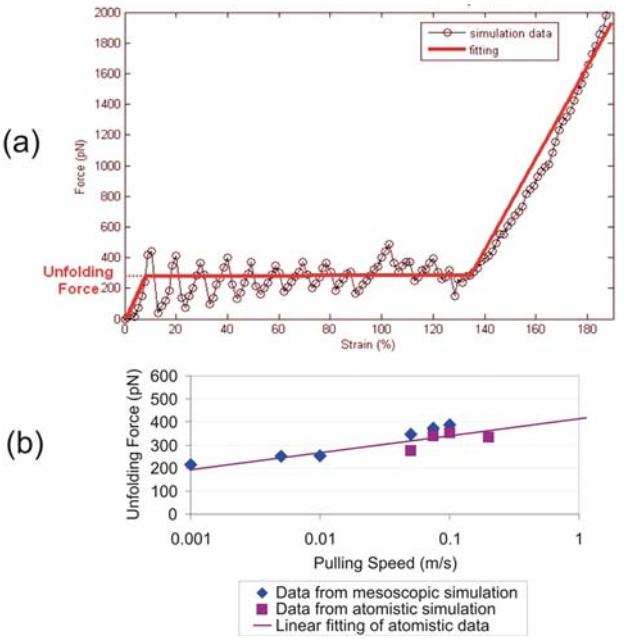


Figure 17-12. Validation of the mesoscopic model by comparison with full atomistic results [94]. Subplot (**a**) shows the entire force-strain curve for a stretching experiment on the 14 beads mesoscopic model of an alpha-helix (with a length of 70.2 Å at a temperature of 300 K). The curve shows the three typical regimes observed in full atomistic simulations: an elasticity regime, an energy dissipation regime which corresponds to the unfolding of the 13 bonds (13 peaks on the curve), and the regime of stretching of the backbone bonds. Subplot (**b**) shows the rate dependence of the unfolding force for both the mesoscopic and atomistic models. The mesoscale model is in very good agreement with the full atomistic simulations so that it validates the fitting of the mesoscopic bond potential

Figure 17-12 presents the validation of our mesoscale model by direct comparison with full atomistic results of the rupture mechanics of an alpha helix protein domain. Figure 17-12a depicts the entire force-strain curve for a stretching experiment on the 14 beads mesoscopic model of an alpha-helix with a length of 70.2 Å, at a temperature of 300 K. The curve shows the three typical regimes observed in full atomistic simulations: an elasticity regime at low strain, an energy dissipation regime which corresponds to the unfolding of the 13 bonds (13 peaks on the curve), and the subsequent regime of stretching of the backbone bonds. The unfolding of the 13 bonds occurs at a steady force which defines the unfolding force. Figure 17-12b reveals the rate dependence of the unfolding force for both the mesoscopic and atomistic models. The mesoscale model is in very good quantitative agreement with the full atomistic simulations. These results validate the fitting of the bond potential, and show that coarse-graining can be used to accurately describe the behavior of protein domains across multiple time-scales, at much reduced computational cost.

In summary, the coarse-grained model enables modeling of the dynamics of large systems over a large range of length- and time-scales. This model enables us to reach time-scales of several microseconds and longer with a quantitative accuracy comparable with full atomistic MD simulations, within several days of computational time (on a single Intel Xeon CPU). In comparison, MD simulations of the dynamical behavior at fractions of microseconds can take weeks and months of computational time (even on a large parallelized simulation setup). This reflects a considerable speedup due to the coarse-graining approach. Potential applications of the coarse-grained alpha-helix protein domain model reviewed here could be studies of length-scale effects on alpha helix strength, elasticity and effects of hierarchical arrangements of alpha-helical based protein domains [94].

17.2.4.6. Case Study: Network Model of Alpha Helices

In this section we review studies of the behavior of a square lattice network of alpha-helical protein filaments, serving as a simplistic model for the nuclear lamin network, as shown in Figure 17-13 [95]. This case study provides another example of how the use of coarse-graining techniques can be used to traverse through several orders of magnitude of length-scales and thereby provide a link between the molecular structure of a single protein domain and the overall behavior of a larger protein network at micrometer length-scales. Figure 17-13a depicts the coarse-graining approach of the AH. Here groups of convolutions are represented by a bead, and thus this model is much coarser than the one discussed in the previous section. (In principle, the formulation discussed above could also be used to study the behavior of protein networks; however, the computational cost would be challenging in systems that approach fractions of micrometers of length-scales.)

Through reducing the degrees of freedom in the coarse-graining approach, this simple model of an alpha-helix enables us to capture the different regimes of deformation at much reduced computational cost. We are now interested in elucidating the behavior of a protein network on the order of length-scales of micrometers. Figure 17-13b displays the overall geometry of the protein network considered here (the overall dimensions that reach 0.4×0.35 μm). We consider a square network structure, defined here to resemble the lamin nuclear envelope (admittedly, this is a simplistic model of this protein network, but is used here to illustrate the general approach of how such a study could be set up and carried out). The particular network structure is motivated by the TEM analysis of the lamin nuclear envelope protein network as shown in Figure 17-13c, which suggests that a square lattice structure of alpha-helical protein filaments is found in cells [96]. We have added a penny-shaped defect inside the protein network to mimic the effect of inhomogeneities of the network architecture. These inhomogeneities can also be seen in the TEM picture shown in Figure 17-13c; three of them are marked in light color to illustrate the concept.

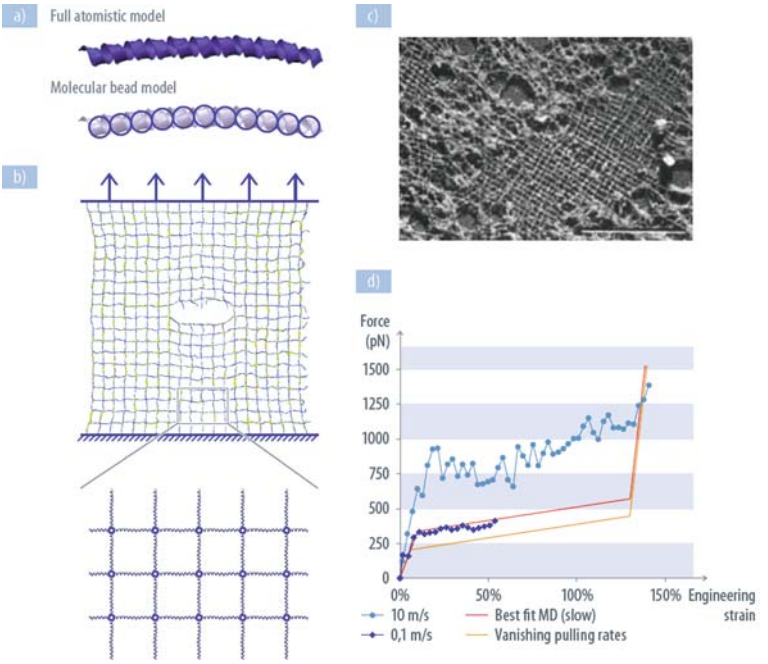


Figure 17-13. Mesoscale protein network model, used here to study the deformation mechanics of a lamin intermediate filament network [95]. Subplot (**a**) shows a schematic of the coarse-graining procedure, from a full atomistic representation arriving at the molecular bead model. Subplot (**b**) depicts a schematic of the coarse grained protein network used in this study with the applied boundary conditions. The size of the network equals to 12×12 nm. Thereby each filament is represented by one alpha helix as shown in the blow-up. A constant strain rate is applied in *y*-direction. We study networks with and without cracks (here an example with crack; the crack represents geometrical flaws/inhomogeneities as they appear in vivo). Subplot (**c**) shows a snapshot of a quasi-regular lamin meshwork as it was observed in experiment (scale bar 5 μm). Figure in subplot (**c**) reprinted from Aebi et al., *Nature*, Copyright © 1986, with permission from MacMillan Publishers Ltd. [198]. Subplot (**d**) depicts force-strain curves from MD simulation results and the approximation used in the mesoscale model. For our mesoscale model we use force values at the AP, which were calculated at vanishing pulling rates in the AR regime (approximately 200 pN). We assume covalent bond rupture of the backbone at forces of 7,800 pN, as calculated with ReaxFF reactive force fields in previous studies (not shown here) [199]

As shown in Figure 17-13a, the network is constructed based on a collection of beads, with a lattice dimension of 10 nm. All pairs of beads (that is, nearest neighbors) in the system interact via a force field description that behaves as shown schematically in Figure 17-13d, capturing the various regimes of deformation as determined from full-atomistic simulation studies. Unlike in the previous case study where a double well potential was used to describe the unfolding of individual convolutions in the alpha helix, here we utilize a multi-linear potential formulations so that the effective force-extension behavior matches that seen in atomistic simulation studies (for further information about this approach, see [74]).

We now investigate how this protein network with a structural flaw behaves under remotely applied tensile deformation. Tensile load is applied by displacing the upper and lower part of the slab in the direction of tensile (mode I) loading. Figure 17-14 displays a sequence of snapshots that illustrate how the protein network undergoes deformation. It can be observed that large portions of the domain undergo large unfolding deformation. The presence of the flaw in the material is tolerated without formation of very large stress concentrations, and large strains of more than 100% strain do not lead to failure of the protein network. Mitigation of these large mechanical strains despite the presence of crack-like defects is due to the particular structure of the alpha-helical protein building block, which provides the basis to its characteristic behavior. Repeated breaking of clusters of H-bonds (with 3-4 H-bonds each, as discussed in Section 17.4.1) provides the mechanistic basis for the network structure to reach large reversible deformation on the order of 150–200% (referred to as "superelasticity", as discussed at the molecular level in [97]) before covalent bonds are stretched and permanent damage occurs in the protein network [97, 98]. This feature enables this network to return to the undeformed perfect structure and shape

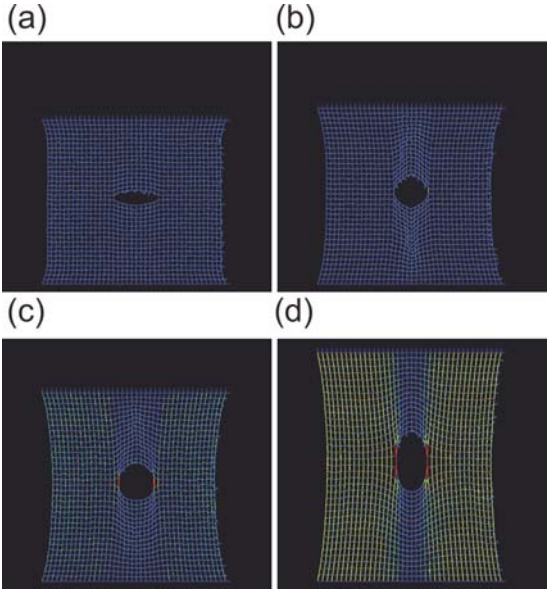


Figure 17-14. Snapshots of the simulation results of tensile deformation of the alpha-helix based protein network, mimicking the lamin intermediate filament structure [95]. It is observed that large portions of the domain undergo large unfolding deformation (subplots (**a**)–(**d**)). The presence of the flaw in the material is tolerated and large strains of more than 100% strain do not lead to failure of the protein network. Thereby, the toleration of these large mechanical strains despite the presence of defects is due to the particular structure of the alpha-helical protein building block, which provides the basis to its characteristic behavior. Repeated breaking of clusters of H-bonds provides the structural basis so that each protein constituent can reach large deformation on the order of 150–200% before covalent bonds are stretched

after the load is released, due to the self-healing property of alpha helix based protein structure because of the underlying H-bonding. Biologically, the significance of this behavior is that it may explain how alpha-helix based protein networks can tolerate defects and structural flaws to mitigate the impact of large deformation of the cell nucleus without failure (to visualize this effect, compare with the imperfect, inhomogeneous structure shown in Figure 17-13c).

The overall shape of the simulated stress-strain curve is in good agreement with experimental results (see Figure 17-15, where experimental data recently published by Fudge et al. on hagfish slime threads [99, 100] is plotted in comparison with the predictions made by our model). In addition to the overall shape of the mechanical signature, the levels of strain are in good agreement as well, albeit there exists some deviation at larger strain levels. Simulation and experiment show a change from the flat to the steep regime at about 100% strain and the inset of rupture at about 150% strain (it is believed that at sliding sets in experimental studies, a feature not yet captured in the multi-scale model). The absolute stress levels diverge due to different geometries of the networks (e.g. length of AHs, and number of parallel AHs).

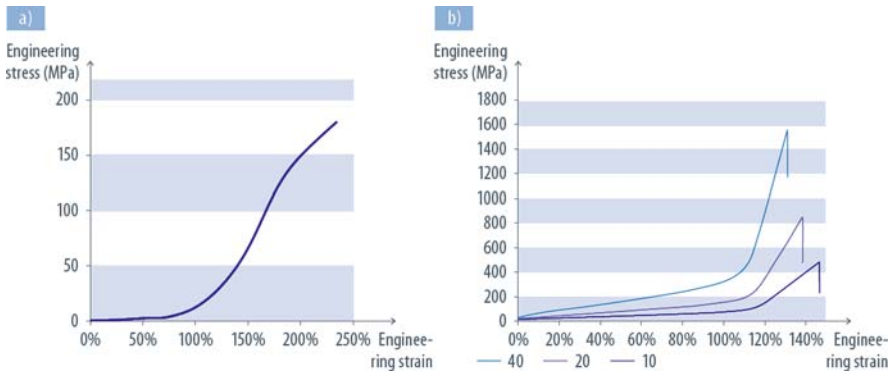


Figure 17-15. Subplot (**a**) shows experimental data on the stress-strain behavior of hagfish slime threads (replotted data from reference [99]). Hagfish slime threads mainly consist of bundles of alpha helical IFs and are thus a good model for comparison with results from mesoscale simulations. Our results show very good agreement with experimental findings regarding the shape of the curve. Both curves have a very flat regime followed by a very steep increase in stress. The change in both curves appears at approximately 100% strain. At about 150% strain flattening sets in, which was suggested as intermolecular sliding. The stress levels mainly depend on the geometry of the bundles/networks as shown in Subplot (**b**). Subplot (**b**) shows our results for different network densities. All networks have a size of 120 × 120 nm. The number indicates how many filaments were arranged in this network in parallel. We observe a strong dependence between the number of parallel strands and the network stiffness as well as the rupture stress. With decreasing rupture forces the rupture strain increases continuously. This is an interesting finding as the stress property itself already represents a force property normalized by the surface area (number of helices)

17.3. THEORETICAL STRENGTH MODELS OF PROTEIN CONSTITUENTS

The strength of biological protein materials is determined by bonds of different strength and how they interplay in the hierarchical structural arrangement. In particular weak H-bonds play a crucial role in defining protein constituents' mechanical strength, and their behavior is critical in formulating bottom up strength models. In the following sections, we review the development of theoretical strength models that emphasize the important role that weak interatomic bonds play in defining the strength properties of biological protein materials.

17.3.1. Strength of a Single Bond

17.3.1.1. Bell's Model: A Force Dependent Dissociation Rate

The Bell's model is a simple and quite popular phenomenological model which describes the frequency of failure of reversible bonds [101]. The concept of "reversibility" thereby means that an individual bond can break under no force if one waits a sufficiently long time, and that it can reform spontaneously. Such bonds may be associated with electrostatic, van der Waals (vdW), or H-bond interactions. The frequency of failure, also called dissociation rate or off rate, is defined as the inverse of the bond lifetime and often used as a concept to discuss the dynamical behavior of such bonds.

Bell's theory explains the force dependence of the off rate and thus shows the significant role of mechanical force in biological chemistry. For instance, this theory can be applied to describe the forced unbinding of biological adhesive contacts such as adhesion of cells to cells. Bell's theory is an extension of the transition state theory for reactions in gases developed by Eyring and others [102]. Inspired also by Zhurkov's work on the kinetic theory of the strength of solids [103], Bell predicted for the first time that the off rate of a reversible bond, which is the inverse of the bond lifetime, increases when subjected to an external force f. Indeed, the rupture of bonds occurs via thermally assisted crossing of an activation barrier E_b which is reduced by $f \cdot x_b$ as the applied force f increases, x_b being the distance between the bound state and the transition state (see Figure 17-16). Thus the Bell off rate expression is

$$k = \omega_0 \exp\left(-\frac{E_b - f \cdot x_b}{k_B \cdot T}\right), \tag{17-8}$$

where ω_0 is a natural vibration frequency and $k_B \cdot T$ the thermal energy. The force $f_0 = E_b/x_b$ represents the force to vanish completely the energy barrier and gives a very rough value of the rupture force.

This conjecture was established long before single molecule experiments were performed. Later, it became very successful especially to describe forced unfolding of biological molecules. Indeed, the model enables one to characterize the bonds, their ruptures and their energy landscape profiles from the fitting with experimental

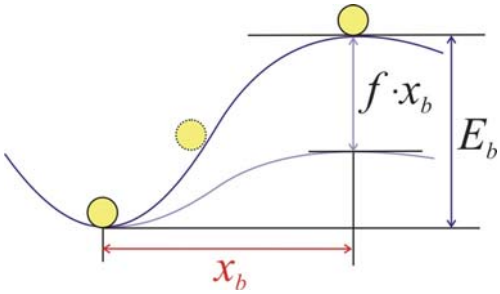


Figure 17-16. Evolution of the energy landscape of a bond subjected to a force. The minimum corresponds to the bound state. The transition state is the peak of the potential and corresponds to the bond rupture. The parameter x_b is the distance between the bound state and the transition state. According Bell's theory, the rupture of a bond occurs via thermally assisted crossing of an activation barrier E_b which is reduced by $f \cdot x_b$ as the applied force f increases

or simulation results. Lastly, as discussed in a following section, this theory leads to the development of new theoretical models to explain more complex molecular bonds such as multiple bonds in parallel.

Although successful, the approach has some limitations which have led to several refinements. A few major refinements are reviewed in more details in the next sections. A major limitation is that the Bell's theory deals only with constant external force and does not explain loading rate dependence of strength. This limitation is important since we usually cannot apply a constant force in experiments due to the very short thermal impulse time compared with the measurement time [104]. Another limitation is the fact that x_b does not depend on the force. In addition, the multidimensional nature of the energy landscape of biomolecules which can lead to multiple unfolding pathways is not taken into account [105].

17.3.1.2. Evans' Extension: A Loading Rate Dependence of Strength

The Evans extension to the Bell model attempts to solve some of the limitations of the Bell's model. It gives a more general relation of the off rate by taking into account force dependent terms such as $x_b(f)$. Moreover it explains for the first time the rate dependence of strength.

The general off rate relation derived by Evans is extended from the Brownian dynamics theory of Kramer. Kramer's theory deals with reactions kinetics in liquids. It gives the thermal noise-driven rate of escape of a particle over a potential barrier using the Fokker-Planck approximation [102, 104]. Evans extended it by including an external pulling force f. Thus he reduced the dimensions of the reaction path to one dimension (coordinate x) and added the potential $-f \cdot x$. Now we briefly present the steps of derivation of the off rate. The time evolution of the probability density of the molecular configuration $p(x,t)$ is described by the Fokker-Planck equation. From the solution, we derive the bond survival probability at time t, referred to $S(t)$, and the mean first passage time, T, which defines the bond lifetime (mean time of passage from the bound state to the unbound state). Thus, we obtain the bond off

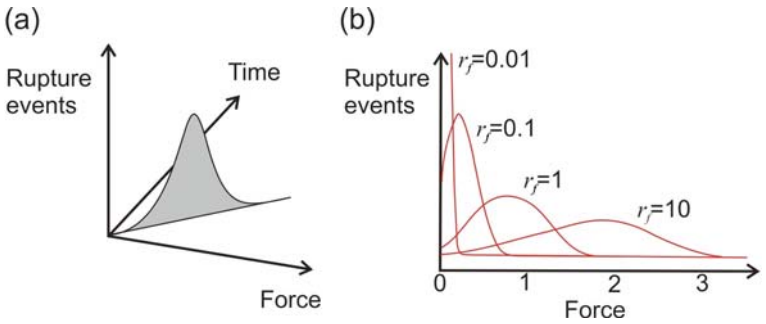


Figure 17-17. Subplot (**a**) illustrates the rupture forces predicted for bonds over time and force in mechanical probe tests [104]. Loaded by a ramp of force, the off rate increases steadily, but the likelihood of bond survival decreases simultaneously. Thus the frequency of failure can reach a maximum at some time equivalent to force; the peak defines the bond strength f^*. Subplot (**b**) shows that rupture force distributions shift with loading rate r_f. Figure adapted based on Ref. [104]

rate as the inverse of the bond lifetime:

$$k = k_0 g(f) \exp\left(\frac{\Delta E_b(f)}{k_B T}\right) \tag{17-9}$$

where k_0 is a prefactor which contains the Arrhenius dependence on barrier energy scaled by a characteristic time constant, $g(f)$ is a function which depends on deformation of energy landscape by external force, and $\Delta E_b(f)$ is the reduction in energy barrier height. This off rate relation is more general compared with Bell's relation. Indeed, it does not use the linear approximation $\Delta E_b(f) = -f \cdot x_b$ for the reduction in energy barrier height. Thus it may be physically more relevant since it allows the positions of transition state and bound state to change under external force.

The second and probably most important contribution of Evans was to demonstrate that the strength of bonds depends crucially on the loading rate. He theoretically showed that, above a critical loading rate, the force of rupture increases logarithmically with the loading rate, and thus explained what was already observed by many experiments. Now, we briefly present the steps of derivation to obtain the rupture force expression at a given loading rate. The force f applied to the bond increases with time t such as $f = r_f \cdot t$ where r_f is the loading rate. Because of thermal fluctuations, the rupture events are stochastic and thus give a rupture force density distribution $p(f)$ (see Figure 17-17a). The probability $p(f) = p(t = f/r_f)$ that the bond breaks at time t is the probability that the bond survives to time t, $S(t)$, multiplied by the bond off rate $k(f)$ which is the instantaneous probability that the bond breaks at time $t = f/r_f$ given that it survives to time t:

$$p(f) = p(t = f/r_f) = k(f) \cdot S(t) \tag{17-10}$$

The bond strength f^* is defined as the most probable rupture force and corresponds to the peak of the force density distribution. Thus we can obtain f^* from the following equation

$$\frac{dp(f)}{df}\Big|_{f=f^*} = 0 \tag{17-11}$$

Provided that x_b and E_b remain constant, one can demonstrate the logarithmic dependence of strength on the loading rate as follows:

$$f^* = \frac{k_B \cdot T}{x_b} \log\left(\frac{r_f \cdot x_b}{k_B \cdot T \cdot k_0}\right) \text{for } r_f > r_c \equiv \frac{k_B \cdot T \cdot k_0}{x_b}, \tag{17-12}$$

where k_0 is the off rate in the absence of force. For smaller loading, $r_f < r_c$, the most probable rupture force is zero. Figure 17-17b illustrates this increase of the most probable rupture force with the loading rate.

In force probe techniques, a transducer is usually moved at constant speed relative to a substrate in which the bond is anchored. The pulling speed v is linked to the loading rate r_f through the stiffness K of the transducer $r_f = K \cdot v$. Thus, this description of rate dependence is very significant for experiments because it enables to get the bond constants k_0 and x_b from a simple linear regression on the force-log(r_f) curve. Moreover it rationalizes the variation among rupture force values obtained from different experimental and simulation techniques which were using different loading rates. However, for a wide range of loading rate we may get a non-linear behavior. An explanation can be that x_b and E_b do not remain constant (change of mechanism, multiple energy barriers), or the elasticity contribution changes with the loading rate.

17.3.1.3. Other Refinements of Bell's Model

Several other attempts have been made to extend the very simple Bell's model. For instance, it has been shown that rebinding can have a great impact on strength [107–109]. Similar to a force dependent x_b term, the existence of a rebinding rate can alter the logarithmic rate dependence of rupture force. In non-equilibrium pulling regime, other models attempt also to explain non-logarithmic rate dependence [109]. Moreover, other extensions try to implement the influence of the transducer stiffness in order to explain the disparities in measured unbinding force among different methods [110]. As a last example of extension, we can mention the existence of models which take into account the energy landscape roughness of bonds [105, 111].

17.3.2. Strength of Complex Molecular Bonds

According Bell's theory, the lifetime of an individual weak bond such as an H-bond is very low. However in biology weak bonds can provide more significant strength by forming arrangement of multiple bonds. In this section, we briefly review theoretical models which study strength of various bond arrangements. A common feature

of these models is that they all implement the Bell's model to describe the off rate of a simple bond at a given time *t*, in other words the probability for the bond to break at time *t* given that the bond survives to time *t*.

17.3.2.1. Multiple Bonds in Parallel

Many attempts have been made to study the strength of multiple parallel bonds [101, 106, 108, 112, 113]. These models try to describe the influence of various parameters on the strength. For instance, interesting parameters can be the potential profile of a simple bond, the number of bonds, the rebinding rate (neglected, constant, Bell based, Boltzmann based, etc.), the constant external force, the loading rate and the stiffness of the transducers. Figure 17-18 presents an example of geometry with a given set of parameters.

The theoretical description of the rupture of parallel bonds is traditionally achieved by one of the two approaches presented below, the deterministic approach and the stochastic approach. We first discuss the deterministic approach. This method describes the time evolution of the mean number of closed bonds *N(t)*

$$\frac{dN}{dt} = -N(t) \cdot k_0 \cdot \exp\left[\frac{f \cdot xb}{k_B \cdot T \cdot N(t)}\right] + k_r \cdot [N_0 - N(t)] \tag{17.13}$$

with N_0 the total number of bonds, f the external force, k_r the rebinding rate of a single broken bond and $k_0 \cdot \exp\left[\frac{f \cdot x_b}{k_B \cdot T \cdot N(t)}\right]$ the Bell's relation for the off rate of a single closed bond. Bell used this approach for a cluster of parallel bonds under constant external force and demonstrated the existence of a critical force above what rupture can happen. Seifert extended it by studying the influence of various parameters such as the loading rate, the rebinding rate and the transducer stiffness. This approach is attractive because it is simple and gives analytical solutions. However, it may not be relevant in some cases such as when we consider the cluster rupture irreversible [112].

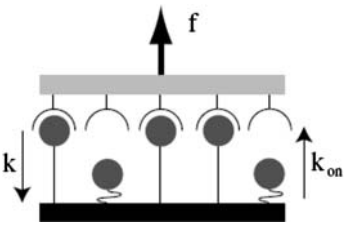


Figure 17-18. Example of a schematic representation of an adhesion cluster under constant force [112]. There are N_0 *(= 5)* bonds in parallel, of which *i (= 3)* are closed and equally share the constant dimensionless force *f*. Single closed bonds rupture with dissociation rate $k = k_0 \exp\left(\frac{f \cdot x_b}{k_B \cdot T \cdot i}\right)$ and single open bonds rebind with force-independent association rate k_{on}. This model has three parameters: cluster size N_0, dimensionless rebinding rate $\gamma = k_{on}/k_0$ and force *f*. Figure reprinted with permission from American Physical Society, from Ref. [112]

Second, we discuss the stochastic approach. This method describes the time evolution of the probability of each possible states. For instance, the probability $p(i)$ that i bonds are closed at time t can be described as follow [112, 113]:

$$\frac{dp_i}{dt} = k_u(i+1) \cdot p_{i+1} + k_r(i-1) \cdot p_{i-1} - [k_u(i) + k_r(i)] \cdot p_i \tag{17.14}$$

with $k_u(i) = i \cdot k_0 \cdot \exp\left[\frac{f \cdot x_b}{k_B \cdot T \cdot i}\right]$ and $k_r(i) = k_r \cdot (N_0 - i)$ the reverse and forward rates between the possible states i. Using the stochastic approach, Erdman and Schwarz studied the bistability of a cluster under no external force applied [114]. Compared with the deterministic approach, the full stochastic approach is more accurate and relevant for a wider range of cases. However the analytical calculation of the solution is more complicated and thus simulation tools such as Monte Carlo methods may be required.

17.3.2.2. Coupled Strength Models

Structures of biological materials can be quite complex and their strength may rely on a variety of physical parameters and phenomena such as biopolymer elasticity and the strength of weak and covalent bonds. In order to obtain an accurate description of the behavior of such materials under mechanical loading, theoretical models may combine different theories. Here, we illustrate this approach through two models which were developed by Rief [104, 105].

First we review a model for biopolymer extensibility [115]. This model combines the Worm Like Chain (WLC) elasticity model with a thermodynamic two-state description extended from Bell's theory. The polymer is made of a series of folded protein domains and is stretched at constant speed. The external force is calculated from the WLC model. From the calculated force and the two-state potential profile, Bell's theory provides the probability of unfolding of each domain. Then Monte Carlo simulation is used to solve the equations and to provide the force versus extension curve corresponding to a particular protein structure. The model can be used to measure the parameters of the two-state potential by fitting the simulation curve of force versus loading rate with the experimental results.

Second, we discuss an elastic bond network model for protein unfolding mechanics [116]. This approach combines an elastic model of a network of bonds with irreversible bond fracture kinetics. The network is subjected to external forces. The elastic model considers bonds as identical springs and calculates the force applied on each bond. Through a Bell theory based model, the unfolding force is predicted. In general, this model may be applied to a variety of different protein structures. It further enables one to study anisotropy and protein unfolding mechanics. Also, it has been suggested that this model can be used with loading rates similar to experimental ones, and thus direct comparisons between experiments and the model are possible [116]. However, it has limitations such as that fracture is irreversible, and that low forces observed in protein rupture cannot be explained.

17.3.2.3. Hierarchical Bell Model

As discussed above, the "original" Bell model (Eq. (17-8)) does not distinguish between a single chemical bond and protein architectures that include several fundamental bonds (e.g. H-bonds). For instance, whether a single H-bond ruptures or if several H-bonds rupture simultaneously is captured in an effective value of E_b. However, this change in mechanism is not explicitly noted in the model given in Eq. (17-8) and subsequent expressions. In order to estimate the strength and the energy landscape of a protein without performing any simulations or experiments and thus to make the model predictive, the Bell model has been extended to explicitly consider the structural hierarchies of the protein structure with the only input parameters being the energy of a H-bond and the rupture distance. The arrangements of H-bonds in protein materials typically represent hierarchical structures, ranging from individual H-bonds at the lowest, atomistic level to a collection of H-bonds at the next higher, molecular protein scale. Figure 17-19 illustrates the approach of representing such hierarchical structures in the Hierarchical Bell Model, and Figure 17-20 shows the application to three model protein domains based on AHs.

In this model, the system breaking force f^{hn} and the energy barrier E_b^{hn} of a system consisting of n hierarchies is given by:

$$f^{hn} = \frac{k_B \cdot T}{x_b \cdot \cos\theta} \cdot \left[\ln\left(\frac{v}{x_b \cdot \omega_0}\right) + \ln\binom{b_n}{k_n} + \sum_{i=2}^{n} b_i \cdot \ln\binom{b_{i-1}}{k_{i-1}} \right] + \frac{\prod_{i=1}^{n} k_i \cdot E_b^0}{x_b \cdot \cos\theta} = f_v + \sum_{i=1}^{n} f_{hi} + f_{h0} \tag{17.15}$$

and

$$E_b^{hn} = \prod_{i=1}^{n} k_i \cdot E_b^0 + k_B \cdot T \cdot \left[\sum_{i=2}^{n} b_i \cdot \ln\binom{b_{i-1}}{k_{i-1}} + \ln\binom{b_n}{k_n} \right]. \tag{17.16}$$

These equations enable one to predict the unfolding force at any pulling speed, once the structural geometry and the energy landscape of a single H-bond is known.

We note that this model only considers H-bonds as structural elements in the definition of alpha helix based protein structures, representing a limitation of this model. Thereby it does not consider hydrophobic effects and other chemical interactions between molecules, which may play a role in influencing the strength properties. This is a limitation of the model; however, there is currently no method to explicitly include these effects in the model and thus this task is thus left to future work. It is expected, however, that the effect of intermolecular adhesion is limited with respect to the prediction of the initial strength values. This

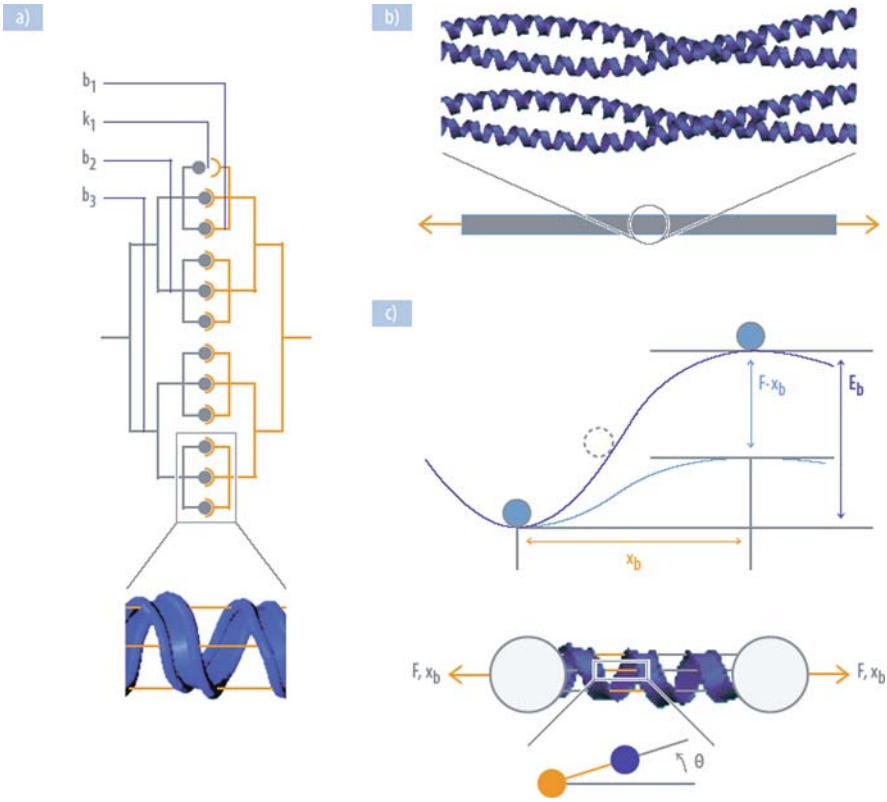


Figure 17-19. Illustration of different hierarchies and their representation in the Hierarchical Bell Theory (subplot (**a**)), as well as representation of the corresponding physical system (subplot (**b**)). Thereby in subplot (**a**) the serial arrangement as it exists in (**b**) is not shown. The inlay in the lower part of subplot (**a**) shows a single alpha helix structure with ≈3 HBs per convolution. The Hierarchical Bell Theory reported here enables one to predict the strength of such hierarchical bond arrangements as a function of the deformation speed. Subplot (**b**) shows the physical system that is represented in the hierarchical model in subplot (**a**). Subplot (**c**): Statistical theory to predict the bond rupture mechanics [117]. The graph depicts the energy as a function of deformation along a deformation variable, along a particular pathway that leads to bond rupture. Here F is the applied force, and x_b is the displacement in the direction of the applied force. In the schematic, three HBs (indicated by the red color) break simultaneously. Thus, x_b corresponds to the lateral displacement that is necessary to overcome the bond breaking distance of a HB

is based on the observation that in alpha helix based coiled-coil protein structures, failure initiates first in the individual alpha helices (thus defining its strength properties) and is later followed by uncoiling of the overall coiled-coil structure [97]. This suggests that the approach taken here by focusing solely on H-bonds is a good approximation for the strength properties of alpha helix based protein domains.

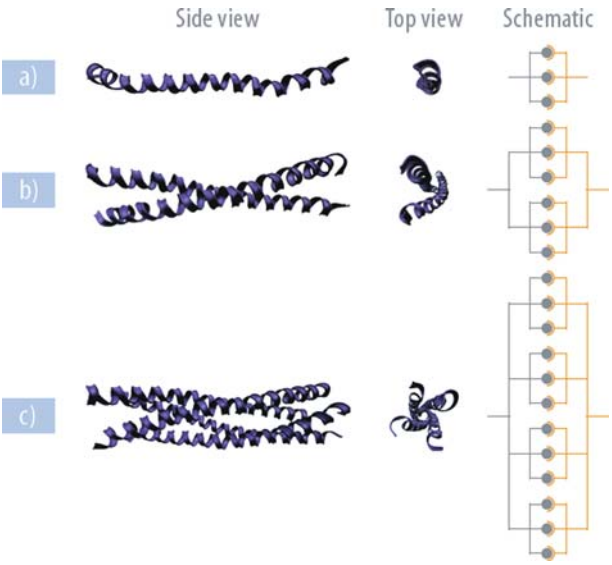


Figure 17-20. Hierarchical Bell model, representation of several protein domains in this model (here for: single AH, coiled-coil, as well as a fourfold coiled-coil protein domain)

17.3.3. Size Effects in H-Bond Clusters

A variety of models for the fracture mechanics of ceramics and metals have been reported over the past decades, involving detailed descriptions of dislocations, plasticity and crack extension mechanisms [11, 12, 117]. However, similar advances for biological protein materials have thus far remained elusive. Recent multi-scale nanomechanical studies have revealed insight into the atomistic mechanisms that govern the mechanical properties of beta-structured materials in protein materials. A rigorous fracture mechanics approach to describe the fundamental bond rupture events in protein materials has been used to explain equilibrium slow-rate mechanical strength limit of beta-structured proteins observed in proteins [118]. In analogy to dislocation nucleation and propagation in ductile metals, rupture processes of H-bonds have been identified as a fundamental unit mechanism of materials failure in protein constituents.

The authors have developed a structure-property-function relationship for beta-sheets, and have shown the size-dependent mechanical and thermodynamical stability of these assemblies [119]. Figure 17-21 shows this scaling law for shear deformation of beta-strand protein domains (subplot a) along with an illustration of the geometry (subplot b), as reported in reference [119]. This result illustrates that beta-strand structures reach their maximum strength at characteristic dimension of H-bond clusters of size between 3 and 4. Interestingly, similar size effects and the existence of a "strongest size" have been observed in metallic polycrystals [120, 121]. Figure 17-22 shows a comparison of the characteristic dimensions

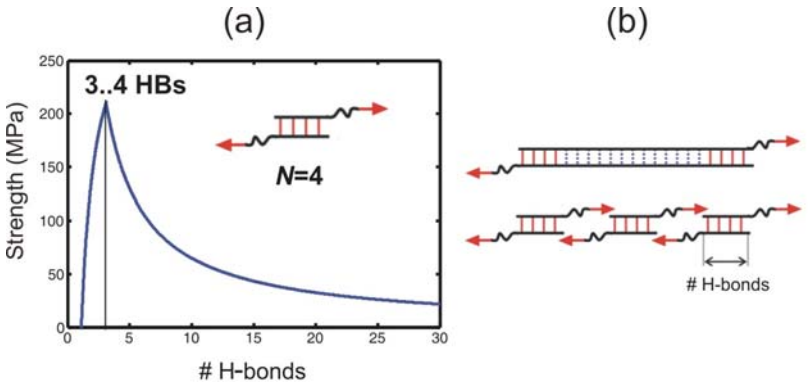


Figure 17-21. Size effects of the shear strength of beta-sheets (the geometry of a beta-sheet is shown in the inlay of subplot (**a**), composed of two polypeptide chains connected through *N* H-bonds; this number of H-bonds has a significant effect on the shear strength of the structure) [119]. Subplot (**a**) depicts the shear strength as a function of the number of H-bonds of a beta-sheet structure as shown in the right half. Subplot (**b**) illustrates the physical significance of this size effect. In the upper plot, only H-bonds at the boundary participate in the rupture process and provide resistance. In the lower plot of this panel, all H-bonds throughout the entire structure contribute to the strength, making the overall structure three times stronger

of H-bond assemblies loaded in parallel in alpha-helices, beta-sheets and beta-helices with the model prediction of the highest shear strength at the characteristic dimension. This analysis suggests that similar size effects may drive the formation and stability of many other basic protein domains found in biological protein materials.

These findings provide ample evidence for the hypothesis that size-effects govern the mechanical behavior of beta-structures at the molecular level, hence motivating further studies in a multi-scale scheme. As discussed in the introduction section, H-bonds have been identified as key chemical interactions that govern the integrity of these structures at the nanoscale and control the large deformation and fracture mechanisms of protein materials. These weak bonds are utilized in biology in geometric confinement and hierarchies to achieve high-strength, robust materials such as spider silk. These findings set the stage for future work in this field, leading to a multi-scale description of the size-dependent properties of hierarchical beta-structures in protein materials.

17.3.4. Asymptotic Strength Model for Alpha Helix Protein Domains

Here we review a constitutive model that characterizes the strength of an alpha-helical (AH) protein domain subjected to tensile deformation, covering more than ten orders of magnitude of time-scales. The model elucidates multiple physical mechanisms of failure in dependence of deformation rate, quantitatively linking atomistic simulation results with experimental strength measurements of

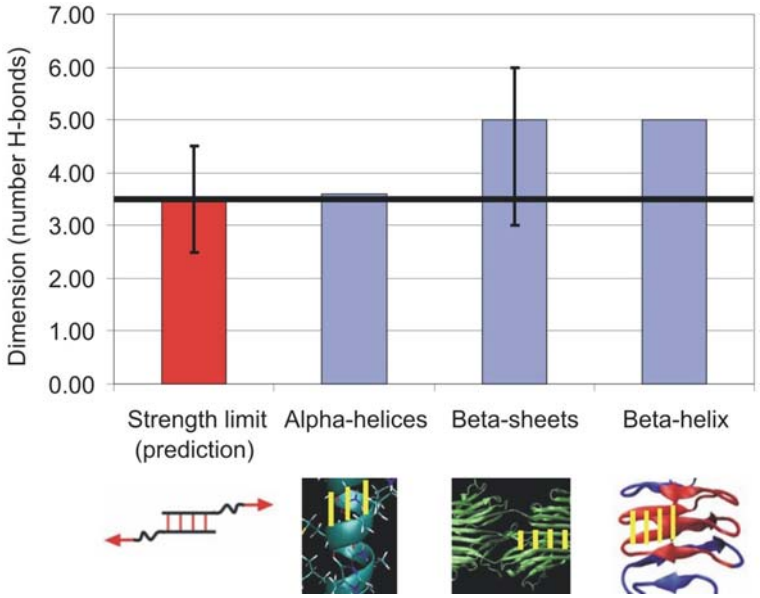


Figure 17-22. Comparison of the characteristic dimensions of H-bond assemblies loaded in parallel in alpha-helices, beta-sheets and beta-helices with the model prediction of the highest shear strength at the characteristic dimension [119]. In conjunction with the theoretical prediction, this plot suggests that geometric confinement may be a universal strategy to create particularly stable protein structures as fundamental material building blocks

AHs. The model provides a description of the strength of AHs based on fundamental physical parameters such as the H-bond energy and the polypeptide's persistence length, showing that strength at high rates is controlled by energetic, nonequilibrium processes and by thermodynamical, equilibrium processes at low rates. This model provides a novel perspective on the strength of protein domains at ultra-slow pulling speeds relevant under physiologic and experimental conditions.

The alpha-helical (AH) protein motif is universally found in structural protein networks and plays an important role in biophysical processes that involve mechanical signals (e.g. mechanotransduction) and provides mechanical integrity to cells [1, 122–124]. For instance, alpha helix rich intermediate filament networks forward mechanical signals from the cell/tissue scale to the DNA [123, 124], aspects that are important for cell mitosis or apoptosis. The mechanical properties of alpha helix protein domains and the link to associated atomic-scale chemical reactions are not only of vital importance in biology, but may also enable the design of synthetic protein materials [125, 126].

The mechanics of AHs plays a crucial role in biology, ranging through disparate time-scales reaching from picoseconds to seconds and more [123, 124, 127, 128]. However, currently there exists no model that describes the mechanical strength

behavior of alpha helix protein domains that considers associated physical mechanisms through this range of time-scales; experiments have been carried out at relatively slow pulling rates, and computer simulations (e.g. molecular dynamics simulations) have been carried out at much faster deformation rates. The results of experiments and computational studies have not yet been integrated. Here we resolve this issue by providing a self-consistent approach that allows us to predict the strength of AHs over more than ten orders of magnitude in time scales, quantitatively linking atomistic simulation results with experimental results, based on fundamental physical parameters that include the energy and geometry of H-bonds and the persistence length of the protein's backbone. The model captures the behavior of AHs from "slow" natural biological processes up to mechanical shock as it appears during accidents or injuries.

17.3.4.1. Modeling and Results

A cartoon of the alpha helix protein and a schematic of the tensile load boundary conditions used to study the rupture mechanism is shown in Figure 17-23. As reported in previous work, molecular dynamics (MD) simulations of AHs in explicit solvent were performed over four orders of magnitudes of pulling speeds (from 0.05 to 100 m/s [39]). The rupture force of the alpha helix structure, identified at the point of breaking of the first H-bonds, is plotted as a function of the protein domain's life time τ in Figure 17-24a. When the system is not in equilibrium, as it is the case

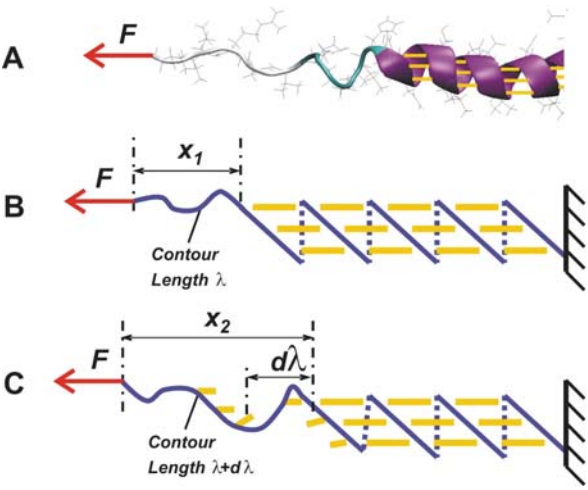


Figure 17-23. Subplot (**a**) depicts the atomistic-scale protein structure of a single alpha helix (AH) from a vimentin coiled-coil dimer [129]. The helical backbone is stabilized by parallel arrangements of hydrogen bonds (HBs, *yellow dashed lines*). Subplots (**b**) and (**c**) show a schematic model system of an alpha helix strained by an external force before and after onset of rupture, showing the process of releasing a segment of backbone polypeptide due to the rupture of HBs, thereby increasing the contour length of the free end entropic chain by $d\lambda$

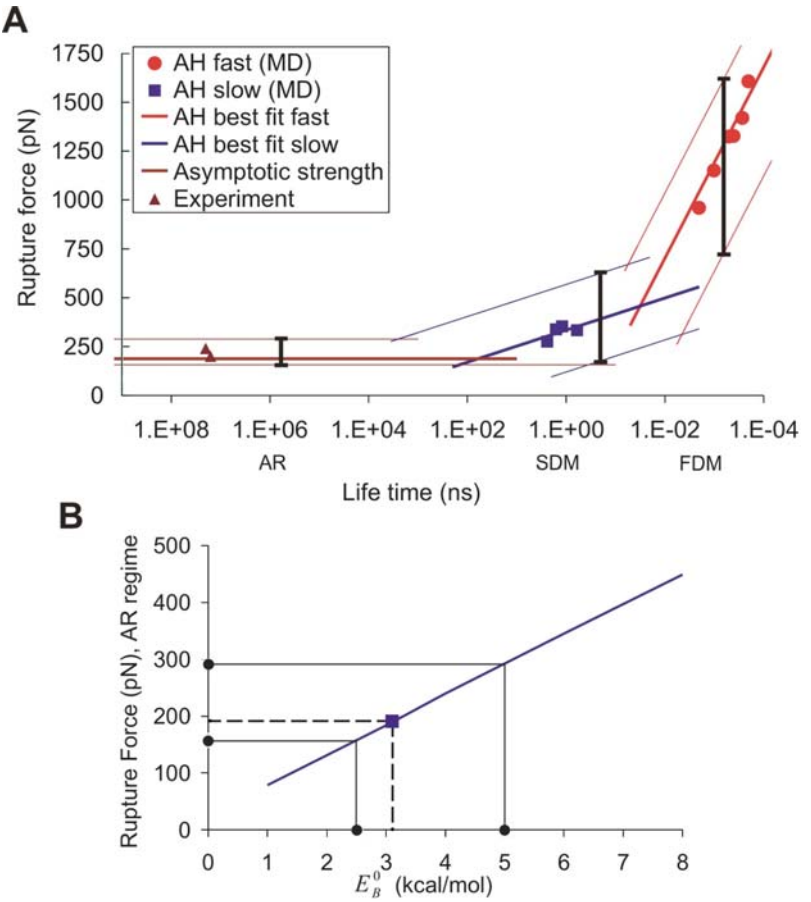


Figure 17-24. Strength model over more than ten orders of magnitudes in time-scale. Subplot (**a**) shows the rupture force versus life time of the AH-system at the onset of failure (=strength properties), including all three regimes over more than ten orders of magnitude of time-scales [129]. MD simulation results (as reported in [39]) suggest a change in mechanism from the fast deformation mode (FDM) to the slow deformation mode (SDM) at increasing the time scales. At approximately 350 pN the effective energy barriers under the applied force in the Bell model are comparable, and therefore mark the transition between FDM and SDM mechanisms. At longer time-scales there is another change in deformation mechanism from the SDM to the asymptotic regime (AR), predicted here at a time scale of approximately 100 ns when $f_{AR} > f_{SDM}$. Experimental results confirm this prediction. Thin lines show the strength behavior for a broad range of HB energy values from 2.5 to 5 kcal/mol (marking error bars for uncertainties in the H-bond energy). Subplot (**b**): Dependence of the critical rupture force on E_B^0, in the AR. The strength of the system near equilibrium conditions (AR) depends linearly on E_B^0 (this parameter determines the energy release rate γ). The specific value ofE_B^0, usually found in a range between 1 and 8 kcal/mol, and varies between different solvent conditions and the specific sequence of the protein domain

for fast deformation rates, the relation between τ and the applied force f can be described by a simple Bell model [130]:

$$\tau = \omega_0^{-1} \cdot \exp\left(\frac{E_b - f \cdot x_b \cdot \cos(\theta)}{k_B \cdot T}\right), \tag{17-17}$$

where E_b is the energy barrier of HB breaking, and x_b is the distance between the equilibrium state and the transition (=rupture) state of the protein domain (note that $v = \Delta x / \Delta t = x_b / \tau$, where v is the externally applied pulling speed). Further, the parameter $\theta \approx 16°$ describes the angle between the applied force f and the orientation of the HBs, k_B is the Boltzmann constant, T is the absolute temperature, and $\omega_0 = 1 \times 10^{13} \mathrm{s}^{-1}$ is the natural frequency of bond vibration. It is noted that in addition to the phenomenological model used here, other stochastic models exist that link time scales and pulling speeds to bond breaking forces; for a description to other models we refer to the literature [104, 131–138]. The force as a function of time scale τ and the energy landscape parameters (ELP, E_b and x_b) is given by

$$f(\tau, E_b, x_B, \theta) = (x_B \cdot \cos(\theta))^{-1} \cdot [E_b - (k_B \cdot T \cdot \ln(\omega_0 \cdot \tau)]\,. \tag{17-18}$$

(for detailed explanations of these equations see [39]). For a given pair of ELPs, Eq. (17-18) leads to a straight line in the f-ln(τ) space. Direct MD simulation studies in explicit water [39] confirm this predicted behavior, however, we observe two distinct regimes, each of which follows the predicted linear logarithmic dependence of the unfolding force with respect to the life time of the structure. The analysis of the atomistic mechanisms of rupture together with the analysis based on Bell's model shows that the two slopes shown in Figure 17-24a correspond to two distinct unfolding mechanisms with two different energy barriers (see Table 1) [39]. In the fast deformation mode (FDM), the observed deformation mechanism and the calculated E_B^{FDM} indicate that single HBs break sequentially, whereas in the slow deformation mode (SDM) 3-4 HBs break simultaneously (3.6 HBs form one alpha-helical convolution, which unfolds as a whole). The sequential breaking of HBs at high pulling speeds (short time scales, FDM) is due to the fact that HB breaking in the protein remains localized. This is because pulling occurs faster than the ability of the protein to mediate HB breaking induced "plastic" deformation. In the SDM regime however, pulling is slow enough so that entire convolutions rupture under the applied force, leading to effectively higher energy barriers for unfolding [39].

At increasing time scales in the SDM the Bell model prediction leads to negative forces, an unphysical prediction. Furthermore, experimental values [139, 140] clearly do not lie on an extension of the slope predicted from the SDM regime, and rather suggests that the f-ln(v) curve approaches an asymptotic zero slope (see Figure 17-2a). Could the Bell model be used to explain this behavior at vanishing pulling rates? Adopting the Bell model to describe this behavior would lead to an increase of x_B (since x_B controls the slope of the f-ln(τ) curve), approaching infinity for slopes approaching zero. It is noted that in other models (e.g. the

microscopic theory [135, 141]) a similar approach has been taken, where the value of x_b is defined as a function of pulling speed (equivalently, the time-scale), leading to a continuous change of the slope of the f-ln(v) curve.

The approach of x_B to extremely large values is, however, unphysical since the transition point x_B can not be larger than the finite contour length of the protein domain. This suggests that another mechanism must determine the protein rupture force. The key to understand this change in mechanism is the realization that at sufficiently long time scales the deformation of the system goes through equilibrium and is no longer controlled by a statistically activated process as described in the Bell model or Eqs. (17-17 and 17-18). Thus the strength does not depend on the time-scale of loading beyond a critical τ_{crit}, and is independent of pulling rate for very long time-scales.

At long time scales $\tau > \tau_{\text{crit}}$ entropic effects that stem from conformational changes of the polypeptide chains are activated and the strength is characterized by a free energy release rate condition, as recently reported in [142] for another class of protein domains. Here we apply this model [142] to alpha helix protein domains. Similar to the Griffith condition used to predict the onset of fracture in crystals [143], the free energy released by freeing polypeptide chains from their geometric confinement in helical convolutions, must equal the energy required to break these HBs. The strength of the alpha helix protein domain is then given by

$$f_{AR} = \frac{k_B T}{4\xi_P}\left[(1-\alpha_{cr})^{-2} + 4\alpha_{cr} - 1\right]. \tag{17-19}$$

with α_{CR} obtained from

$$G(\alpha_{cr}) = \frac{k_B T}{4\xi_P}\left[\alpha_{cr}\cdot(1-\alpha_{cr})^{-2} - (1-\alpha_{cr})^{-1} + 2\alpha_{cr}^2 + 1\right] \stackrel{!}{=} \gamma. \tag{17-20}$$

Hereby α equals to the ratio of the end-to-end length of the free chain to its contour length $\alpha = x/\lambda$ (see Figure 17-1 for definition of variables), equivalent to mechanical stretch. The parameter γ describes the HB energy stored per unit length of AH,

$$\gamma = \frac{E_B^0 \cdot n}{l} = \frac{E_B^0}{L_0}, \tag{17-21}$$

with l as the unit length of one convolution and n as the number of HBs per unit length.

We calculate $L_0 = 0.145$ nm/HB by measuring the length of the entire protein ($L_x = 6.9$ nm) and dividing it by the number of existing HBs ($n = 47$). This is in good agreement with results in the literature [122] where $L_x = 0.15$ nm/HB (calculated from $l = 0.54$ nm and $n = 3.6$ HBs per convolution). The strength regime described by Eq. (17-19) is referred to as asymptotic regime (AR).

Combing all three mechanisms (FDM, SDM, AR), the strength of a alpha helix domain is:

$$F(\tau; x_b^{\mathrm{FDM}}, E_b^{\mathrm{FDM}}, x_b^{\mathrm{SDM}}, E_b^{\mathrm{SDM}}, \theta, \xi_p, \gamma) = \max \left\{ \begin{array}{c} f_{\mathrm{FDM}}\left(\tau; x_b^{\mathrm{FDM}}, E_b^{\mathrm{FDM}}, \theta\right) \\ f_{\mathrm{SDM}}\left(\tau; x_b^{\mathrm{SDM}}, E_b^{\mathrm{SDM}}, \theta\right) \\ f_{\mathrm{AR}}\left(\xi_p, \gamma\left(E_b^0\right)\right) \end{array} \right\}. \tag{17-22}$$

The functions f_{FMD} and f_{SMD} can be calculated from Eq. (17-22), f_{AR} can be calculated from Eqs. (17-19, 17-20, 17-21). We estimate E_B^0 from the MD simulation in the FDM where the 3.6 H-bonds in one convolution break simultaneously, thus $E_b^0 = E_b^{\mathrm{SDM}}/3.6 = 3.1$ kcal/mol, thus $\gamma = 2.1$ kcal/mol/Å (this provides a direct link between SDM and AR). This value of E_b^0 is in good agreement with earlier experimental and simulation results [144], where E_b^0 was reported to be 3–6 kcal/mol. We choose the persistence length of a polypeptide chain as suggested from both experiment and theory to $\xi_P = 4$ Å [145]. Based solely on these two parameters, E_b^0 and ξ_P, the force in the AR is calculated to ≈190 pN. The AR regime is reached at time scales of 100 ns (or equivalently, at pulling speeds $v <$ 0.001 m/s), when $f_{AR} > f_{SDM}$. The strength value of f_{AR} is plotted in Figure 17-24b as a function of the HB energy E_b^0.

The model given in Eq. (17-22) is validated through quantitative comparison with experimental results. Experimental results of stretching and breaking single alpha helix domains [139, 140] (with a length of less than 100 Å) report forces between 140 and 240 pN during unfolding. Figure 17-24a summarizes the described regimes and shows a quantitative comparison between the model prediction and MD simulation results as well as experimental results. In addition to the values used in this study that were based on earlier MD results, an envelope curve for E_b^0 ranging from 2.5 to 5 kcal/mol is included to illustrate how the predictions change under variations of the energy of HBs. We note that other experimental results [145–149] (not shown in Figure 17-2) that consider alpha helix spectrin repeats lay slightly below the predicted force range, on the order of 50 pN, which would require extremely low values of E_b^0 ≈1 kcal/mol. A possible explanation for this behavior could be the difference in the observed unfolding mechanism, which is the unfolding of the anti-parallel coiled-coil repeat instead of rupture of individual HBs of a alpha helix domain. For instance, in one of the studies x_b was estimated to be 15 Å [145–149], which is ten times higher than the x_b for a single HB thus suggesting an alternative mechanism.

17.3.4.2. Summary and Discussion

The most important feature of the model reviewed in this section is the development of a constitutive model (Eq. (17-22)) that describes the strength properties of alpha helix protein domains over more than 10 orders of magnitudes of time-scales. Up until now such a model has not been reported, and to the best of our knowledge this model is the first to quantitatively link MD simulation results [39] and experimental alpha helix strength values [139, 140] in a simple physical model as shown in Figure 17-24a. An important feature of the model reported in Eq. (17-22) is that it only includes basic parameters of the protein structure, that is, the HB energy and

geometry, as well as persistence length. The strength properties of the alpha helix protein domain, a universally found biological protein structure, is controlled by different mechanisms at distinct time scales, with strong strengthening under faster rates (shorter time-scales).

According to this model, the strength at very slow pulling rates is controlled by entropic effects of the freed polypeptide backbone and not from a continuously changing energy barrier that moves along the reaction coordinate x_B as suggested in the microscopic theory. Changes in the reaction coordinate x_B are only observed at relatively fast pulling rates, where it can be directly linked to changes in the physical mechanism of rupture (that is, the change from FDM to SDM). This study could motivate new experiments, in particular those that would provide a systematic variation of deformation rates to probe the transitions between the regimes described here.

17.4. COMPLEMENTARY EXPERIMENTAL METHODS

The science of the mechanical behavior of protein materials relies on complementary experimental investigations. Experimental methods, used either to explore the structures and properties or for the synthesis of novel materials, are of great importance to balance theoretical studies and are crucial to develop a full understanding of these materials, throughout all scales.

17.4.1. Structural Characterization

The most important experimental techniques in conjunction with atomistic modeling of protein materials are x-ray diffraction and Nuclear Magnetic Resonance (NMR) analysis, which provide detailed information on the crystallographic structure and chemical bonding at the molecular level, including the three-dimensional atomistic geometry of protein molecules. Furthermore, atomistic modeling can also feed back to assist structural characterization of protein materials, by providing information about conformational stability, structural dynamics, as well as the interactions of proteins with solvents. The structure of many proteins, elucidated using x-ray diffraction or NMR experiments, has been deposited in the Protein Data Bank [150], providing a rich source of three-dimensional structures that can be used as the starting point for molecular simulation studies, or for structural analyses of the geometry of protein components. Experimental techniques based on spectroscopy and microscopy techniques have also been developed to monitor the dynamics of chemical bonds and molecules, as illustrated for instance in a recent study of H-bond dynamics in a protein structure [151].

17.4.2. Manipulation and Mechanical Testing

Recent advances in experimental techniques and nanotechnologies provide more possibilities to observe, explore, and even mechanically manipulate the functional components of biological protein materials at the fundamental molecular level. The so-called force spectroscopy techniques at molecular level, such as atomic force

microscopy, optical tweezers and nanoindentation, allow one to measure mechanical properties of proteins, macromolecules, and even individual chemical bonds. These studies can provide valuable insight to analyze the molecular mechanisms associated with mechanical failure of protein domains and larger-scale filaments [26–31, 152].

The Atomic Force Microscope (AFM) consists of a micro-cantilever with a sharp tip at its end, which is used to scan the structure and properties of materials through interatomic interactions such as van der Waals, electrostatic, or capillary forces. The mechanical signature of proteins and other single biomolecules can be obtained by AFM, where the biomolecule (for example, an individual protein domain) is attached to a surface and manipulated by the tip that pulls the molecule at constant force or constant pulling speed (the attachment to surfaces can be achieved by utilizing molecules that display strong binding affinity to protein molecules and to the surface, serving as a molecular link). AFM experiments on protein structures often lead to a characteristic saw-tooth shaped force-displacement profile, and could be linked to sequential unfolding of certain domains in the protein (see Figure 17-8, where such behavior is shown as obtained from a MD simulation study). The Worm Like Chain model (WLC) [153, 154] is frequently used to describe the entropic elasticity of these domains (at least in part of the overall deformation regime), as discussed above.

Optical tweezers apply forces on molecules that result from focused laser beam, a concept called optical trapping. The force can catch, hold and move dielectric particles ranging from nanometers to micrometers, at force levels of up to several hundred pN with pN resolution. This technique is very precise, flexible and damage-free, so has been widely used in the fields of cell biology and biophysics, for example to unravel the mechanism of DNA condensation [155] and many other studies (see, e.g. an overview article [156]). These single-molecular level techniques provide extremely useful information to guide and validate atomistic-based multi-scale computer simulation.

In nanoindentation, a small tip is pressed into the material and the load-displacement relation is recorded. With the help of elasticity theory, the indentation results are analyzed and provide hardness and elastic modulus of the materials. Because of the small size of indentation tip, small-scale, highly localized properties of materials can be measured. Using this method, spatially dependent mechanical properties from bone, teeth to soft tissues have been measured successfully [157], with high spatial and force resolution.

We refer the reader to other articles regarding details of these nanomechanical experimental approaches (see, e.g. [27–29, 31, 158–161]). A selection of experimental techniques is summarized in Figure 17-2, illustrating the overlap with multi-scale simulation methods. Since these advances in experimental methods now enable one to probe time- and length-scales that are also directly accessible to large-scale atomistic based simulation, the combination of experiment and simulation might lead to a particularly fruitful interaction. This is particularly promising since the kind of information obtained from experiment and simulation might be complementary.

17.4.3. Synthesis Methods for Hierarchical Materials

Materials science studies of nanoscale and biological materials benefit significantly from conventional synthesis methods such as lithography, catalytic vapor deposition, and templated growth. Materials synthesized using these techniques range from crystals (e.g. metals, ceramics), thin films (e.g. coatings, electronic devices), to nanostructures of various kinds (e.g. carbon nanotubes, nanowires). These materials are widely used as the functional elements in nanoscale devices. For example, the chemical synthesis of carbon nanotubes was used as cantilever beam in high-frequency nano-resonators [162], and nano/micro-channels created through lithography methods form the basis of microfluidic devices [163].

However, many of the methods that provided great success for the synthesis of "conventional" nanostructures (e.g. CNTs, nanowires, etc.) face difficulties for the synthesis of *hierarchical* nanostructures, since the absence of precise manipulation methods at multiple levels restricts the controlled buildup of such structures. Therefore, self-assembly processes inspired from biological protein materials represent great opportunities for this purpose [164]. Through self-assembly processes, it is possible to form hierarchical nanostructures through interactions such as hydrogen bonding, metal coordination, hydrophobic interactions, van der Waals and/or electrostatic forces. Following this self-assembly approach, programmable protein or nanowires arrays have been constructed [165]. However, self-assembly processes, either static or dynamical, can be very complex. Currently there clearly exists a lack of understanding of fundamental underlying mechanisms [164]. To successfully apply the self-assembly concept for the design and synthesis of hierarchical materials synthesis, we must understand the mechanism of self-assembly, in terms of the constitutive functional elements, in particular their properties and interactions. To facilitate the engineering process and the design of *de novo* hierarchical structures, we must also develop rational design concepts of materials based on our understanding of structure-property-processing relationships.

17.5. *DE NOVO* DESIGN OF BIOINSPIRED AND BIOMIMETIC NANOMATERIALS

The search for materials possessing contrasting properties, such as high strength and high toughness, has driven the design of many alloys and composite materials. However, most synthetic materials are seen to lie on the so-called "banana curve" in regard to these properties, thus possessing one property at the cost of the other (see Figure 17-25a). Biological structural materials, on the other hand, in particular bone [166], nacre [167, 168] and dentin [169] (which are found in the endo- and exoskeleton systems of animals), are stiff and tough simultaneously (see Figure 17-25). Though composed of protein and mineral, they are almost as tough as the protein phase and as stiff as the mineral phase. To introduce a representative example among biological structural materials, we provide a brief review of the structure of cortical bone and an outline of how its contrasting properties and its hierarchical structure are intimately linked. Bone has excellent toughness and

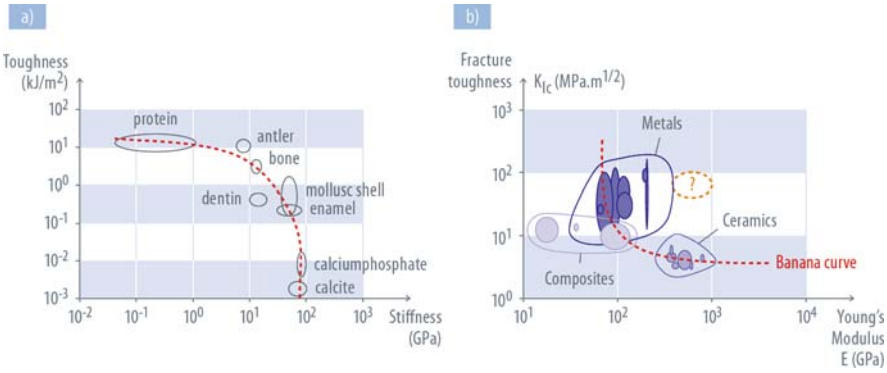


Figure 17-25. (**a**) Toughness *versus* stiffness for a number of biological materials (figure based on the data compilation in reference [200]). Biological composites, such as antler, dentin, bone and enamel combine the "good" properties of the protein and mineral components and are typically both stiff and tough. Generally speaking, the stiffness increases and the toughness decreases with mineral content from antler to dentin/bone and enamel [201]; (**b**) Toughness *versus* stiffness for metals, alloys and ceramics, lie on the "banana-curve", an inverse relation between increasing toughness and decreasing stiffness. The circular region shows the property region of high toughness and stiffness that may be accessible through designing bio-inspired composite structures

fatigue strength properties while being light weight and stiff for load-carrying. Its structure is composed of 7 levels of hierarchical arrangements [166]. This is thus an ideal material that may provide the basis for mimicking concepts to achieve high toughness.

Bone is a composite of organic and inorganic constituents: 30% bone is organic; of which 90–95% is collagen, and the remainder consists of non-collagenous proteins. The remaining 70% of bone is made up of the inorganic mineral hydroxyapatite, which includes calcium phosphate, calcium carbonate, calcium fluoride, calcium hydroxide and citrate [3, 170, 171]. At the nanoscale, bundles of collagen molecules are arranged in fibrils, which are twisted in a coil, making a collagen fiber. The inorganic component is predominantly crystalline. The crystals are platelets or rods, about 8–15 Å thick, 20–40 Å wide and 200–400 Å long and arranged in a regular array at the nanoscale (Figure 17-26a). These mineralized collagen fibers form into planar arrangements called lamellae (3–7 μm wide). These sheets (lamellae) of mineralized collagen fibers wrap in concentric layers around a central bone canal to form osteons. Osteons appear like cylinders with 200–250 μm in diameter, running parallel to the long axis of the bone. Figure 17–26a shows these levels of hierarchy in bone from the nanoscale up to the macroscale.

Bone is known for its combination of superior mechanical properties compared to both of its constituents. Simple reinforced composite models for elastic properties do not give correct results for bone, showing us that the explicit consideration of the hierarchical structure is important [170]. A major property of bone is its increased fracture resistance and toughness, despite being highly mineralized. This has been attributed to distinct deformation mechanisms on different length scales in bone. At

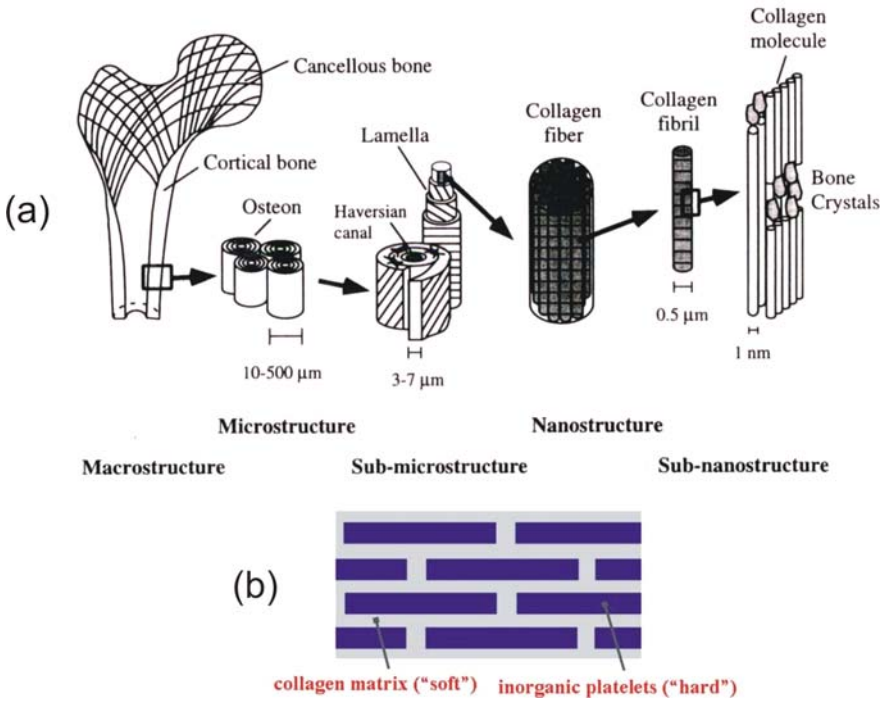


Figure 17-26. Subplot (**a**) shows hierarchical structural organization of bone: (i) cortical and cancellous bone (different types of bone); (ii) osteons; (iii) lamellae; (iv) collagen fiber assemblies of collagen fibrils; (v) bone mineral crystals, collagen molecules, and non-collagenous proteins. Figure reprinted from [166]. Copyright © 1998, with permission from Elsevier. Subplot (**b**) shows two-dimensional, simple schematic model of hard-soft phases nanocomposite based on the ultrastructure of bone. In bone, the inorganic platelets are 2–4 nm thick and up to 100 nm long embedded in a collagen-rich protein matrix

the micron length scale, where the bone structure consists of osteons, this has been attributed to two mechanisms (a) crack bridging, and (b) microcracking [172]. It is hypothesized that microcracks tend to originate around osteons due to debonding at osteon-matrix interface or osteon pull-out. The presence of microcracks in the wake of a crack have been shown to result in the residual opening of the crack tip, and a redistribution of stresses in the crack tip region, which reduces the crack extension force and increase the toughness of the material. Crack bridging in the wake of a crack has also been proposed as a crack tip shielding mechanism. Crack bridging involves formation of unbroken regions that span the crack in the wake of the crack tip and act to resist crack opening. Such bridging can results from un-cracked ligaments and intact collagen fibrils. On the nanometer length scale, bone toughness has been attributed to the flaw tolerance of size of mineral platelets [173]. The nanostructure of bone is shown in Figure 17-26b, consisting of mineral platelets arranged in a staggered pattern in a collagen matrix. This structural motif of long embedded mineral platelets in a soft matrix is seen in other biological materials too (e.g. nacre),

and the commonality of this structural motif suggests some intrinsic properties in the design that improve mechanical properties. A mechanism has been proposed by Gao, Fratzl et al. [46], whereby under tensile loading, staggered mineral platelets carry tensile load and the protein matrix transfers the load between mineral crystals via shear. In this model, it has been showed that the nanoscale width of the mineral platelets embedded in the collagen matrix is such that the material becomes insensitive to crack like flaws at this length scale (approximately tens of nanometers) and fails under tension at the theoretical strength for a perfect crystal [173]. The optimum aspect ratio (height/width) of these platelets can be obtained by assuming that protein and mineral fail at the same time. However, some shortcomings of this simplified model are that it does not take into account the complex non-stoichiometric chemistry at mineral-protein interfaces [174, 175] and size limitations of minerals owing to chemical boundary conditions. The modular domain nature of the organic matrix at different scales, and its stepwise unfolding has been proposed as a mechanism for the intrinsic toughness of the matrix [16, 176], illustrating the hierarchical nature of the matrix collagen itself.

This brief review of the physics of toughness and fracture strength of bone show the importance of hierarchical levels in optimizing mechanical properties. Mechanisms that operate in at least two very distinct length scales are seen to play a prominent role in ensuring the large toughness of bone.

17.5.1. Development of Bioinspired Metallic Nanocomposites

The nanostructural motif (Figure 17-26b) and concept of hierarchical arrangement can be transferred to conventional metal-metal or metal-ceramic nano-composites. Such materials would be light-weight, very strong and energy dissipative. A study of the applicability and transferability of the key design concepts that provide bone with its strength and toughness could for instance be achieved through computational design experiments carried out via molecular dynamics simulation. The use of molecular dynamics provides a detailed description of the atomistic deformation processes under loading at the nanoscale and may serve as a tool to enable the nanoengineering of hierarchical bioinspired materials.

A model nanocomposite material based on the design geometry of bone has been studied using a Ni-Al nanostructure and a modified EAM/EAM alloy potential. This approach has been used to design a composite structure that features a maximized flow stress (results shown in Figure 17-27) [177]. For the Ni-Al nanostructure, in which Ni represents the harder metal, is used as platelet reinforcements in a soft Al matrix. Systematic studies of size variations have shown the existence of a "strongest size" for the platelet size on flow strength under tensile loading [177]. The dominant contribution to plasticity switches over from interfacial sliding and bulk plasticity at this length-scale [177]. Similarly, a model EAM/EAM alloy with Cu metal as matrix material with a modified harder Cu EAM potential as platelets has been used for geometrical parameter studies for strength under tensile deformation and toughness under shock loading of the nanocomposite. The primary deformation

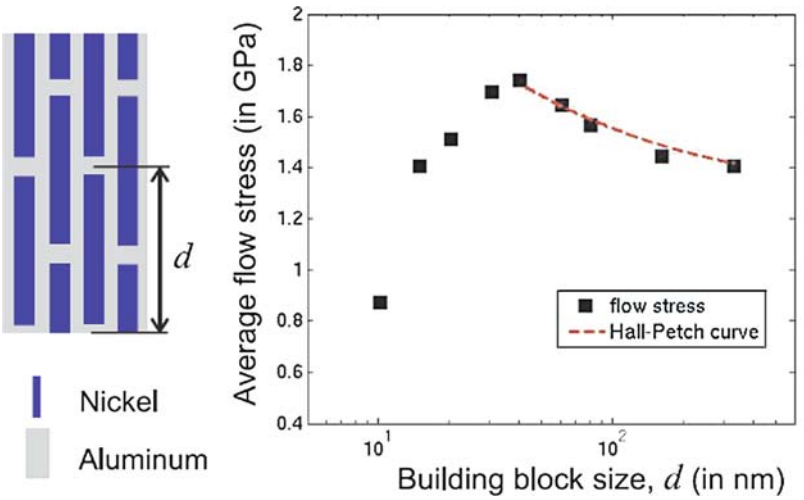


Figure 17-27. Size dependence of the average flow stress (=strength) of a bioinspired metallic nanocomposite, illustrating the concept of a "strongest size" at a characteristic length-scale of material nanostructure [44, 45]. The increase in strength scales very well according to the Hall-Petch (HP) relationship, and the reduction of strength at small building block sizes below approximately 50 nm represents an inverse Hall-Petch regime (IHP)

modes are observed to change over the length scales and geometrical arrangements of the nanocomposite. The possible deformation modes under observation were bulk and platelet plasticity, platelet-matrix decohesion, and interfacial sliding. Low temperature phenomena have been studied, thus discounting diffusional creep mechanisms.

17.5.2. Nanostructure Design Effects Under Tensile and Shock Loading

Atomistic simulations have been carried out that revealed which types of shape, confinement/length-scales and spacing of second phase platelets produces the maximum strength, and what changes in deformation mechanisms are observed due to these geometric changes. Specifically, the effect of varying various geometric parameters across their range have been studied, and parameters have been identified that have significant effect on strength [178]. This approach has been applied to explore the yield mechanics of a nanocomposite with hard platelet inclusions embedded in a soft ductile metal matrix (in a model system of Ni-Al). The main finding of this large-scale atomistic simulation study is that there exists an optimal size of the platelet inclusions that leads to maximum flow stress. The dominant contribution to plasticity is seen to switch over from bulk plasticity to interfacial sliding at this length-scale.

The result is significant in two ways: (i), it shows that the large-deformation mechanical properties can be tuned and dissipation optimized by tailoring the nanostructure, and (ii), it shows that there exists an optimal length scale at which the flow

stress and thus dissipation reach a maximum. The critical geometric parameters for maximum flow stress at large deformation have been determined. The studies provide mechanistic insight into the nanoscopic stress distribution and its relation to the toughening mechanism. This result is significant since it reveals that not only the size, but also the details of the platelet arrangement are significant in determining the overall larger-scale mechanical behavior. The reinforcement platelet offset from one row to another turns out to be a particularly important parameter. For little to no offset, there is no significant strengthening effect (Figure 17-28). For platelet offset larger than one fourth, the flow stress reaches a plateau value and does not increase further. Studies of the effect of platelet shape, as in size aspect ratio, have revealed that elongated platelets in the direction of loading axis provide an increased flow stress (Figure 17-28). However, beyond a critical aspect ratio of elongated platelets, there is no further improvement in strength. On the other hand, keeping the aspect ratio fixed and changing shape of platelets to elliptical, triangular and others shows virtually no effect on the flow stress (Figure 17-28).

The model system of stiff particles embedded in a soft matrix with similar nanoscopic structural features as bone has also been tested under shockwave

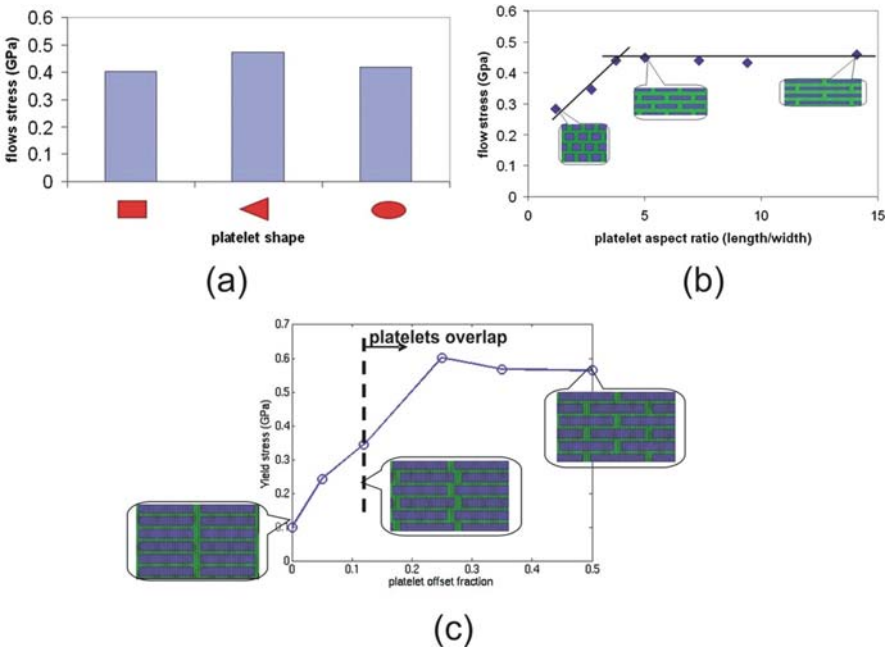


Figure 17-28. Flow stress under tensile loading of bioinspired metallic nanocomposite; (**a**) shows no significant effect of reinforcing platelet phase particle shape on flow stress with fixed volume fraction and aspect ratio; (**b**) shows increase in flow stress as a function of platelet aspect ratio until a critical value, keeping volume fraction and platelet spacing the same. This shows elongated platelets lead to higher flow stresses; (**c**) shows large effect on flow stress of platelet offset from one row to another. For little to no offset, there is no significant strengthening effect. For platelet offset larger than a 1/4th, the flow stress reaches a plateau value

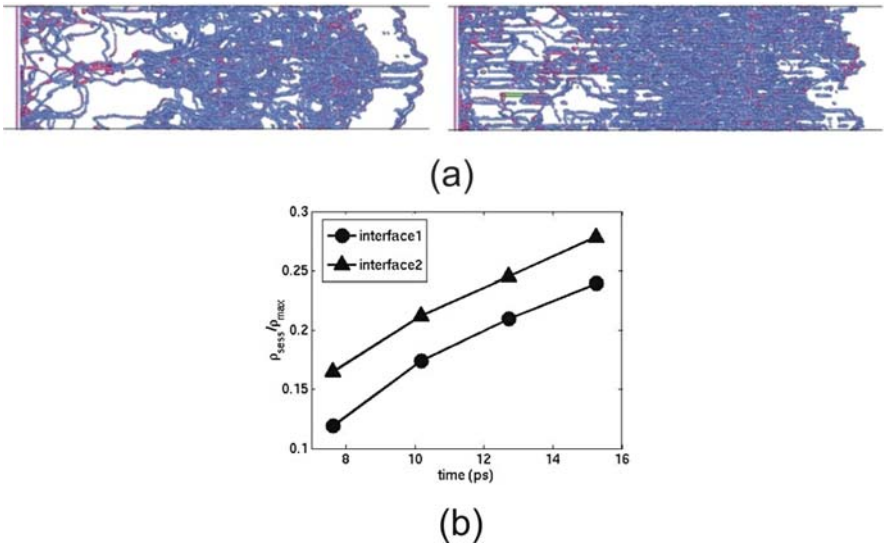


Figure 17-29. Response to shock loading of bioinspired metallic nanocomposites of varying interfacial adhesion between matrix and platelet phases; (**a**) shows dislocation core density for "strong" (interface1) and "weak" (interface2) interfacial strengths. We find that interfacial sliding between the composite's constituents is a major source of plasticity under shock loading, as seen in the horizontal dislocation core lines at the matrix-platelet interfaces for the "weak" interface case; (**b**) shows larger fraction of sessile dislocation activity for the "weak" interface case too pointing towards increased hardening. These results point towards controlling the interfacial strength to design a material with larger shock absorption

loading using large scale molecular dynamics. It is found that the geometric arrangement and the specific length scales of design elements at nanoscale does not have a significant effect on shock dissipation, in contrast to the case of tensile loading where the nanostructural length scales strongly influence the mechanical properties. Interfacial sliding between the composite's constituents is a major source of plasticity under shock loading [179]. Thus controlling the interfacial strength can be used to design a material with larger shock absorption (Figure 17-29).

These results, obtained by using large-scale MD simulations in the spirit of an engineering design approach, could guide laboratory processing techniques in the design of novel nanocomposites.

17.5.3. Outlook and Opportunities

The design attempts to mimic biological structures into synthetic materials can extend beyond optimizing contrasting mechanical properties and improving robustness. Self-organization at various length scales can be mimicked to obtain reversible structures that assemble or disassemble in complex fashions under stimuli [180, 181]. In particular, one of the challenges facing technological advances in developing novel micro- and nano-devices is the effect of miniaturization on thermal management. Miniaturization and higher integration and closer packing of components in electronic devices lead to high-density, point load heat sources at sub

100 nm scales that can lead to failure of NEMS/MEMS devices, significantly reducing their reliability. In these applications, the use of self-assembling heat dissipating structures could possible eliminate the need for the a priori knowledge of the location of heat sources, and enable an adaptive, low weight, self-repairing network. Self-assembling branched fluidic structures mimicking vascular networks of real living tissue were already used to optimize heat management efficiency [182]. Recent nanotechnology based efforts have validated extremely high thermal conductivity ($\approx$ 3,000–6,000 W/mK) in materials such as carbon nanotubes [183] and graphene [184]. However without integration, these nanostructures themselves can hardly be applied to macroscale and feature a lack of controllability. Through controllable fabrication of hierarchical structures [185] or combining with bio-inspired self-assembly strategies based on these conventional high thermal conductivity materials such as nanowires, nanotubes [186] and polymeric inter-links, hierarchical highly dissipative structures can be created for the use in NEMS/MEMS devices, laser diodes and many other applications.

17.6. DISCUSSION AND CONCLUSION

Over the last centuries, engineers have developed quantitative analytic and synthetic understanding of how to create complex structures out of a diverse range of constituents, at various scales (machines, buildings, airplanes, nuclear reactors and many others). Research in the area of material properties of biological materials will extend our ability to carry out structural engineering, as used for buildings or bridges today, to the ultimate scale – nanoscale, and may be a vital component of the realization of nanotechnology through the merger of material and structure [117] (Figure 17-30). The merits of polymers, membranes and other nanostructures could not be utilized sufficiently unless the signal or motion at atomistic scale can be represented though macroscopic functions. Materiomics, the systematic analysis of biological protein materials within the materials science paradigm, may provide a new approach in engineering to create materials that have low impact on environment and energy utilization but are highly effective and functional (e.g. mechanomutable, controllable, changeable).

A better understanding of the mechanics of biological and natural materials, integrated within complex technological systems will make it possible to combine living and non-living environments to develop sustainable technologies. New materials technologies such as protein-based materials produced by recombinant DNA techniques represent new frontiers in materials design and synthesis [187, 188]. These questions have high impact in the understanding and design of environmentally friendly technologies and may enhance the quality of life of millions of people, through advances in the medical sciences as well as through improvements of the living environment. A currently pressing question is the development of new technologies to address the energy problem. Advances may be possible by utilization of bacteria to produce and process fuel from crops, by enabling the synthesis of materials at reduced processing temperature or through fabricating artificial light-harvesting networks.

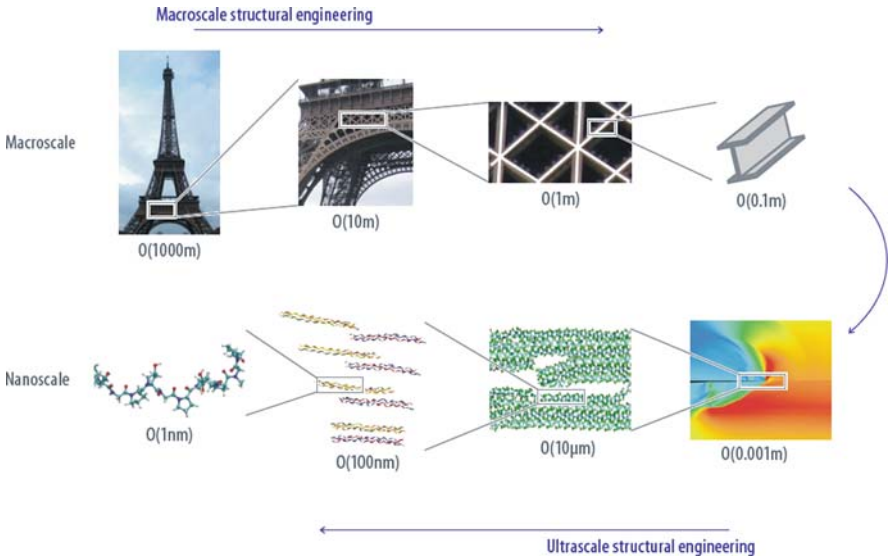


Figure 17-30. Merger of structure and material in engineering design. The long term impact of this work is that it could extend our ability to perform structural engineering at macroscale, to the ultimate scale, the nanoscale. Opening the structural material scale as design space for new material development may open great possibilities for development of robust, adaptive, active and smart materials

Nanoscience and nanotechnology enable us to make structures with unique material properties at the ultimate scale (self assembly, recombinant DNA, utilization of motor proteins for nano-machines and many others). This will perhaps lead to novel complex structural and multi-functional materials, designed from nano to macro. The theoretical progress in understanding hierarchical biological materials will facilitate the use of an extended physical space, through the use of multiple hierarchies, in an efficient and controlled manner, that is, lead to a bottom-up structural design on the sub-macroscopic scale, instead of trial-and-error approaches. For example, the extended design space might serve as mean to realize new physical realities that are not accessible to a single scale, such as material synthesis at moderate temperatures, or fault tolerant hierarchical assembly pathways [189], which enable biological systems to overcome the limitations to particular chemical bonds (soft) and chemical elements (organic) present under natural conditions [190]. The increased understanding of the hierarchical design laws might further enable the development and application of new organic and organic-inorganic multi-featured composites (such as assemblies of carbon nanotubes and proteins or polymer-protein composites [191–193]), which will mainly consist of chemical elements that appear in our environment in an almost unlimited amount (C, H, N, O, S). These materials might consequently help to solve our energy and resource problems (e.g. fossil resources, iron etc.), and allow us to manufacture nano-materials and devices, which will be produced in the future by techniques like recombinant DNA [187, 194, 195] or peptide self-assembly [188, 196, 197], techniques where the bor-

ders between materials, structures and machines vanish. An exciting aspect of the materials science of biological materials is that it is interdisciplinary, by nature. Performing research in this field thus often means to overcome barriers between scientific disciplines and to develop strategies that enable us to communicate concepts more efficiently. Structures in universities and research institutions may have to be modified to facilitate such investigations. To investigate through the materiomics approach towards engineering bio-inspired materials or biological nanomachines, the organization of functional elements at different hierarchical levels, their communications and dynamics must be understood well and held under accurate control. The integration of computational approaches in the design process may play a crucial role in developing these scientific and technological advances.

Applications of these new materials and structures are new biomaterials, new polymers, new composites, engineered spider silk, new scaffolding tissues, improved understanding of cell-ECM interactions, cell mechanics, hierarchical structures and self-assembly. In addition to the long-term impact in biology, bioengineering and medicine, this research may eventually contribute to our theoretical understanding of how structural features at different scales interact with one another. This may transform engineering approaches not only for materials applications, but also in manufacturing, transportation or designs of networks and further the development of the technological approaches.

ACKNOWLEDGEMENTS

This research was supported by the Army Research Office (ARO), grant number W911NF-06-1-0291 (program officer Dr. Bruce LaMattina), the Solomon Buchsbaum AT&T Research Fund, as well as a National Science Foundation CAREER Award (CMMI-0642545, program officer Dr. Jimmy Hsia). Further support from the Air Force Office of Scientific Research (AFOSR), Office of Naval Research (ONR) and DARPA is greatly acknowledged.

REFERENCES

1. Alberts, B., et al., *Molecular Biology of the Cell*. 2002, New York: Taylor & Francis.
2. Astbury, W.T. and A. Street, *X-ray studies of the structures of hair, wool and related fibres. I. General*. Transactions of the Royal Society of London A, 1931. **230**: 75–101.
3. Weiner, S. and H.D. Wagner, *The material bone: Structure mechanical function relations*. Annual Review of Materials Science, 1998. **28**: 271–298.
4. Currey, J.D., *Bones: Structure and Mechanics*. 2002, Princeton, NJ: Princeton University Press.
5. Lakes, R., *Materials with structural hierarchy*. Nature, 1993. **361**(6412): 511–515.
6. Wegst, U.G.K. and M.F. Ashby, *The mechanical efficiency of natural materials*. Philosophical Magazine, 2004. **84**(21): 2167–2181.
7. Vincent, J.F.V., *Structural Biomaterials*, Edited by Anonymous. 1990, Princeton, NJ: Princeton University Press, p. 244.
8. Fratzl, P. and R. Weinkamer, *Nature's hierarchical materials*. Progress in Materials Science, 2007. **52**(8): 1263–1334.

9. Aizenberg, J., et al., *Skeleton of Euplectella sp.: Structural hierarchy from the nanoscale to the macroscale*. Science, 2005. **309**(5732): 275–278.
10. Courtney, T.H., *Mechanical Behavior of Materials*. 1990, New York, USA: McGraw-Hill.
11. Broberg, K.B., *Cracks and Fracture*. 1990, London: Academic Press.
12. Hirth, J.P. and J. Lothe, *Theory of Dislocations*. 1982, New York: Wiley-Interscience.
13. Fraser, P. and Bickmore, W., *Nuclear organization of the genome and the potential for gene regulation*. Nature, 2007. **447**(7143): 413–417.
14. Engler, A.J., et al., *Matrix elasticity directs stem cell lineage specification*. Cell, 2006. **126**(4): 677–689.
15. Buehler, M.J., *Atomistic and continuum modeling of mechanical properties of collagen: Elasticity, fracture and self-assembly*. Journal of Materials Research, 2006. **21**(8): 1947–1961.
16. Buehler, M.J., *Nature designs tough collagen: Explaining the nanostructure of collagen fibrils*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(33): 12285–12290.
17. Fratzl, P., et al., *Structure and mechanical quality of the collagen-mineral nano-composite in bone*. Journal of Materials Chemistry, 2004. **14**: 2115–2123.
18. An, K.N., Y.L. Sun, and Z.P. Luo, *Flexibility of type I collagen and mechanical property of connective tissue*. Biorheology, 2004. **41**(3–4): 239–246.
19. Ramachandran, G.N. and G. Kartha, *Structure of collagen*. Nature, 1955. **176**: 593–595.
20. Doyle, J., *Rules of engagement*. Nature, 2007. **446**: 860.
21. Kitano, H., *Computational systems biology*. Nature, 2002. **420**(6912): 206–210.
22. Kitano, H., *Systems biology: A brief overview*. Science, 2002. **295**(5560): 1662–1664.
23. Gautieri, A., S. Uzel, S. Vesentini, A. Redaelli, M.J. Buehler, *Molecular and mesoscale disease mechanisms of Osteogenesis Imperfecta.* Biophysical Journal, 2009. **97**(3): 857–865.
24. Suresh, S., et al., *Connections between single-cell biomechanics and human disease states: gastrointestinal cancer and malaria*. Acta Biomaterialia, 2005. **1**(1): 15–30.
25. Cross, S.E., et al., *Nanomechanical analysis of cells from cancer patients*. Nature Nanotechnology, 2007. **2**: 780–783.
26. Smith, B.L., et al., *Molecular mechanistic origin of the toughness of natural adhesives, fibres and composites*. Nature, 1999. **399**(6738): 761–763.
27. Prater, C.B., H.J. Butt, and P.K. Hansma, *Atomic force microscopy*. Nature, 1990. **345**(6278): 839–840.
28. Sun, Y.L., et al., *Stretching type II collagen with optical tweezers*. Journal of Biomechanics, 2004. **37**(11): 1665–1669.
29. Dao, M., C.T. Lim, and S. Suresh, *Mechanics of the human red blood cell deformed by optical tweezers*. Journal of the Mechanics and Physics of Solids, 2003. **51**(11–12): 2259–2280.
30. Tai, K., F.J. Ulm, and C. Ortiz, *Nanogranular origins of the strength of bone*. Nano Letters, 2006. **11**: 2520–2525
31. Lim, C.T., et al., *Experimental techniques for single cell and single molecule biomechanics*. Materials Science & Engineering C-Biomimetic and Supramolecular Systems, 2006. **26**(8): 1278–1288.
32. Goddard, W.A., *A perspective of materials modeling* in *Handbook of Materials Modeling*, Edited by S. Yip. 2006, Berlin: Springer.
33. Csete, M.E. and J.C. Doyle, *Reverse engineering of biological complexity*. Science, 2002. **295**(5560): 1664.
34. Stelling, J., et al., *Robustness of cellular functions*. Cell, 2004. **118**(6): 675–685.
35. Aizenberg, J., et al., *Skeleton of Euplectella sp.: Structural hierarchy from the nanoscale to the macroscale*. American Association for the Advancement of Science, 2005. **309**: 275–278.

36. Currey, J.D., *Materials science: Hierarchies in biomineral structures*. Science, 2005. **309**: 253–254.
37. Woesz, A., et al., *Micromechanical properties of biological silica in skeletons of deep-sea sponges*. Journal of Materials Research, 2006. **21**(8): 2069.
38. Horn, J., N. Nafpliotis, and D.E. Goldberg, *A niched Pareto genetic algorithm for multiobjective optimization*. Evolutionary Computation, 1994. IEEE World Congress on Computational Intelligence, Proceedings of the First IEEE Conference on 1994, pp. 82–87.
39. Ackbarow, T., et al., *Hierarchies, multiple energy barriers and robustness govern the fracture mechanics of alpha-helical and beta-sheet protein domains*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**: 16410–16415
40. Ackbarow, T. and M.J. Buehler, *Hierarchical coexistence of universality and diversity controls robustness and multi-functionality in protein materials*. Theoretical and Computational Nanoscience, 2008. **5**(7): 1193–1204.
41. Buehler, M.J., S. Keten, and T. Ackbarow, *Theoretical and computational hierarchical nanomechanics of protein materials: Deformation and fracture*. Progress in Materials Science, 2008. **53**(8): 1101–1241.
42. Tang, Z., et al., Nanostructured artificial nacre. Nature Materials, 2003. **2**(6): 413–418.
43. Vashishta, P., R.K. Kalia, and A. Nakano, *Large-scale atomistic simulations of dynamic fracture*. Computing in Science and Engineering, 1999. **1** (5): 56–65.
44. Rountree, C.L., et al., *Atomistic aspects of crack propagation in brittle materials: Multimillion atom molecular dynamics simulations*. Annual Review of Materials Research, 2002. **32**: 377–400.
45. Buehler, M.J., *Atomistic Modeling of Materials Failure*. 2008, New York: Springer.
46. Buehler, M.J. and H.J. Gao, *Dynamical fracture instabilities due to local hyperelasticity at crack tips*. Nature, 2006. **439**(7074): 307–310.
47. Buehler, M.J., F.F. Abraham, and H. Gao, *Hyperelasticity governs dynamic fracture at a critical length scale*. Nature, 2003. **426**: 141–146.
48. Buehler, M.J. and H. Gao, *Ultra large scale atomistic simulations of dynamic fracture* in *Handbook of Theoretical and Computational Nanotechnology*, Edited by W. Schommers and A. Rieth. 2006, Stevenson Ranch, CA: American Scientific Publishers (ASP).
49. Buehler, M.J., A.C.T.v. Duin, and W.A. Goddard, *Multi-paradigm modeling of dynamical crack propagation in silicon using the ReaxFF reactive force field*. Physical Review Letters, 2006. **96**(9): 095505.
50. Buehler, M.J., et al., *Threshold crack speed controls dynamical fracture of silicon single crystals*. Physical Review Letters, 2007. **99**: 165502
51. Wang, W., et al., *Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions*. Annual Review of Biophysics and Biomolecular Structure, 2001. **30**: 211–243.
52. Mackerell, A.D., *Empirical force fields for biological macromolecules: Overview and issues*. Journal of Computational Chemistry, 2004. **25**(13): 1584–1604.
53. Deniz, A.A., S. Mukhopadhyay, and E.A. Lemke, *Single-molecule biophysics: at the interface of biology, physics and chemistry*. Journal of the Royal Society Interface, 2008. **5**(18): 15–45.
54. Scheraga, H.A., M. Khalili, and A. Liwo, *Protein-folding dynamics: Overview of molecular simulation techniques*. Annual Review of Physical Chemistry, 2007. **58**: 57–83.
55. Van der Spoel, D., et al., *GROMACS: Fast, flexible, and free*. Journal of Computational Chemistry, 2005. **26**(16): 1701–1718.
56. Nelson, M.T., et al., *NAMD: A parallel, object oriented molecular dynamics program*. International Journal of Supercomputer Applications and High Performance Computing, 1996. **10**(4): 251–268.

57. Ponder, J. and D. Case, *Force fields for protein simulations*. Protein Simulations, 2003. **66**: 27–85.
58. MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. Journal of Physical Chemistry B, 1998. **102**(18): 3586–3616.
59. Mayo, S.L., B.D. Olafson, and W.A. Goddard, *Dreiding – A generic force-field for molecular simulations*. Journal of Physical Chemistry, 1990. **94**(26): 8897–8909.
60. Rappe, A.K., et al., *Uff, a full periodic-table force-field for molecular mechanics and molecular-dynamics simulations*. Journal of the American Chemical Society, 1992. **114**(25): 10024–10035.
61. Pearlman, D.A., et al., *Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules*. Computer Physics Communications, 1995. **91**(1–3): 1–41.
62. Gao, H., *A theory of local limiting speed in dynamic fracture*. Journal of the Mechanics and Physics of Solids, 1996. **44**(9): 1453–1474.
63. Duin, A.C.T.v., et al., *ReaxFF: A reactive force field for hydrocarbons*. Journal of Physical Chemistry A, 2001. **105**: 9396–9409.
64. Brenner, D.W., et al., *A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons*. Journal of Physics-Condensed Matter, 2002. **14**(4): 783–802.
65. Stuart, S.J., A.B. Tutein, and J.A. Harrison, *A reactive potential for hydrocarbons with intermolecular interactions*. Journal of Chemical Physics, 2000. **112**(14): 6472–6486.
66. Strachan, A., et al., *Shock waves in high-energy materials: The initial chemical events in nitramine RDX*. Physical Review Letters, 2003. **91**(9): 098301-1–098301-4.
67. Nielson, K.D., et al., *Development of the ReaxFF reactive force field for describing transition metal catalyzed reactions, with application to the initial stages of the catalytic formation of carbon nanotubes*. Journal of Physical Chemistry A, 2005. **109**: 49.
68. Duin, A.C.T.v., et al., *ReaxFF SiO: Reactive force field for silicon and silicon oxide systems*. Journal of Physical Chemistry A, 2003. **107**: 3803–3811.
69. Han, S.S., et al., *Optimization and application of lithium parameters for the reactive force field, ReaxFF*. Journal of Physical Chemistry A, 2005. **109**(20): 4575–4582.
70. Chenoweth, K., et al., *Simulations on the thermal decomposition of a poly(dimethylsiloxane) polymer using the ReaxFF reactive force field*. Journal of the American Chemical Society, 2005. **127**(19): 7192–7202.
71. Strachan, A., et al., *Thermal decomposition of RDX from reactive molecular dynamics*. Journal of Chemical Physics, 2005. **122**(5): 054502
72. Cheung, S., et al., *ReaxFF(MgH) reactive force field for magnesium hydride systems*. Journal of Physical Chemistry A, 2005. **109**(5): 851–859.
73. Chenoweth, K., et al., *Development and application of a ReaxFF reactive force field for oxidative dehydrogenation on vanadium oxide catalysts*. Journal of Physical Chemistry C, 2005. **112**: 14645–14654.
74. Buehler, M.J., *Hierarchical chemo-nanomechanics of stretching protein molecules: Entropic elasticity, protein unfolding and molecular fracture*. Journal of Mechanics of Materials and Structures, 2007. **2**(6): 1019–1057.
75. Datta, D., A.C.T.v. Duin, and W.A. Goddard, *Extending ReaxFF to Biomacromolecules*. Unpublished, 2005.
76. Buehler, M.J., et al., *The Computational Materials Design Facility (CMDF): A powerful framework for multiparadigm multi-scale simulations*. Materials Research Society Proceedings, 2006. **894**: LL3.8.
77. Tozzini, V., *Coarse-grained models for proteins*. Current Opinion in Structural Biology, 2005. **15**(2): 144–150.

78. Tirion, M., *Large amplitude elastic motions in proteins from a single-parameter, atomic analysis*. Physical Review Letters, 1996. **77**(9): 1905–1908.
79. Haliloglu, T., I. Bahar, and B. Erman, *Gaussian dynamics of folded proteins*. Physical Review Letters, 1997. **79**(16): 3090–3093.
80. Hayward, S. and N. Go, *Collective variable description of native protein dynamics*. Annual Review of Physical Chemistry, 1995. **46**: 223–250.
81. West, D.K., et al., *Mechanical resistance of proteins explained using simple molecular models*. Biophysical Journal, 2006. **90**(1): 287–297.
82. Dietz, H. and M. Rief, *Elastic bond network model for protein unfolding mechanics*. Physical Review Letters, 2008. **1**(9): 098101-1–098101-4.
83. Sulkowska, J.I. and M. Cieplak, *Mechanical stretching of proteins – a theoretical survey of the Protein Data Bank*. Journal of Physics-Condensed Matter, 2007. **19**(28): 283201.
84. Bathe, M., *A finite element framework for computation of protein normal modes and mechanical response*. Proteins-Structure Function and Bioinformatics, 2008. **70**(4): 1595–1609.
85. Bahar, I. and R. Jernigan, *Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation*. Journal of Molecular Biology, 1997. **266**(1): 195–214.
86. Nguyen, H. and C. Hall, *Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(46): 16180–16185.
87. Nguyen, H. and C. Hall, *Spontaneous fibril formation by polyalanines: Discontinuous molecular dynamics simulations*. Journal of the American Chemical Society, 2006. **128**(6): 1890–1901.
88. Arkhipov, A., P. L. Freddolino, et al., *Coarse-grained molecular dynamics simulations of a rotating bacterial flagellum.* Biophysical Journal, 2006. **91**(12): 4589–4597.
89. Buehler, M.J., *Nature designs tough collagen: Explaining the nanostructure of collagen fibrils*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(33): 12285–12290.
90. Buehler, M., *Molecular nanomechanics of nascent bone: fibrillar toughening by mineralization*. Nanotechnology, 2007. **18**(29): 295102.
91. Chen, J., W. Im, and C. Brooks, *Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field*. Journal of the American Chemical Society, 2006. **128**(11): 3728–3736.
92. Chen, J., C. Brooks, and J. Khandogin, *Recent advances in implicit solvent-based methods for biomolecular simulations*. Current Opinion in Structural Biology, 2008. **18**(2): 140–148.
93. Roux, B. and T. Simonson, *Implicit solvent models*. Biophysical Chemistry, 1999. **78**(1–2): 1–20.
94. Bertaud, J., Z. Qin, M.J. Buehler, *Atomistically informed mesoscale model of alpha-helical protein domains,* I International Journal for Multiscale Computational Engineering, 2009. **7**(3): 237–250.
95. Ackbarow, T., D. Sen, C. Thaulow, and M.J. Buehler, *Alpha-helical protein networks are self protective and flaw tolerant,* PLoS ONE, 2009. **4**(6): e6015.
96. Aebi, U., et al., *The nuclear lamina is a meshwork of intermediate-type filaments*. Nature, 1986. **323**(6088): 560–564.
97. Ackbarow, T. and M.J. Buehler, *Superelasticity, energy dissipation and strain hardening of vimentin coiled-coil intermediate filaments: Atomistic and continuum studies*. Journal of Materials Science, 2007 **42**(21): 8771–8787.
98. Buehler, M.J. and S. Keten, *Elasticity, strength and resilience: A comparative study on mechanical signatures of α-helix, β-sheet and tropocollagen domains*. Nano Research, 2008. **1**(1): 63–71.
99. Fudge, D.S., et al., *The mechanical properties of hydrated intermediate filaments: Insights from hagfish slime threads*. Biophysical Journal, 2003. **85**(3): 2015–2027.

100. Fudge, D.S. and J.M. Gosline, *Molecular design of the alpha-keratin composite: insights from a matrix-free model, hagfish slime threads*. Proceedings of the Royal Society of London Series B-Biological Sciences, 2004. **271**(1536): 291–299.
101. Bell, G.I., *Models for the specific adhesion of cells to cells*. Science, 1978. **200**(4342): 618–627.
102. Hanggi, P., P. Talkner, and M. Borkovec, *Reaction-rate theory: Fifty years after Kramers*. Review of Modern Physics, 1990. **62**(2): 251–341.
103. Zhurkov, S.N., *Kinetic concept of the strength of solids.* International Journal of Fracture Mechanics, 1965. **1**: 311–323.
104. Evans, E. and K. Ritchie, *Dynamic strength of molecular adhesion bonds*. Biophysical Journal, 1997. **72**(4): 1541–1555.
105. Hyeon, C. and D. Thirumalai, *Measuring the energy landscape roughness and the transition state location of biomolecules using single molecule mechanical unfolding experiments*. Journal of Physics, Condensed Matter, 2007. **19**(11): 113101.
106. Seifert, U., *Rupture of multiple parallel molecular bonds under dynamic loading*. Physical Review Letters, 2000. **84**(12): 2750–2753.
107. Seifert, U., *Dynamic strength of adhesion molecules: Role of rebinding and self-consistent rates*. Europhysics Letters, 2002. **58**(5): 792–798.
108. Evans, E., *Probing the relation between force-lifetime-and chemistry in single molecular bonds*. Annual Reviews in Biophysics and Biomolecular Structure, 2001. **30**(1): 105–128.
109. Hummer, G. and A. Szabo, *Kinetics from nonequilibrium single-molecule pulling experiments*. Biophysical Journal, 2003. **85**(1): 5–15.
110. Walton, E.B., S. Lee, and K.J. Van Vliet, *Extending Bell's model: How force transducer stiffness alters measured unbinding forces and kinetics of molecular complexes*. Biophysical Journal, 2008. **94**(7): 2621.
111. Zwanzig, R., *Diffusion in a rough potential*. Proceedings of the National Academy of Sciences of the United States of America, 1988. **85**(7): 2029–2030.
112. Erdmann, T. and U.S. Schwarz, *Stability of adhesion clusters under constant force*. Physical Review Letters, 2004. **92**(10): 108102.
113. Erdmann, T. and U.S. Schwarz, *Bistability of cell-matrix adhesions resulting from nonlinear receptor-ligand dynamics*. Biophysical Journal, 2006. **91**(6): L60.
114. Erdmann, T. and U.S. Schwarz, *Stability of adhesion clusters under constant force*. Physical Review Letters, 2004. **92**(10): 4.
115. Rief, M., J.M. Fernandez, and H.E. Gaub, *Elastically coupled two-level systems as a model for biopolymer extensibility*. Physical Review Letters, 1998. **81**(21): 4764–4767.
116. Dietz, H. and M. Rief, *Elastic bond network model for protein unfolding mechanics*. Physical Review Letters, 2008. **100**(9): 98101.
117. Buehler, M.J. and T. Ackbarow, *Fracture mechanics of protein materials*. Materials Today, 2007. **10**(9): 46–58.
118. Keten, S. and M.J. Buehler, *Asymptotic strength limit of hydrogen bond assemblies in proteins at vanishing pulling rates*. Physical Review Letters, 2008. **100**: 198301.
119. Keten, S. and M.J. Buehler, *Geometric confinement governs the rupture strength of H-bond assemblies at a critical length scale*. Nano Letters, 2008. **8**(2): 743–748.
120. Yip, S., *The strongest size*. Nature, 1998. **391**: 532–533.
121. Wolf, D., et al., *Deformation mechanism and inverse Hall-Petch behavior in nanocrystalline materials*. Zeitschrift für Metallkunde, 2003. **94**: 1052–1061.
122. Gruber, M. and A.N. Lupas, *Historical review: Another 50th anniversary – new periodicities in coiled coils*. Trends in Biochemical Sciences, 2003. **28**(12): 679–685.
123. Moir, R.D. and T.P. Spann, *The structure and function of nuclear lamins: implications for disease*. Cellular and Molecular Life Sciences, 2001. **58**(12–13): 1748–1757.

124. Wilson, K.L., M.S. Zastrow, and K.K. Lee, *Lamins and disease: Insights into nuclear infrastructure*. Cell, 2001. **104**(5): 647–650.
125. Bryson, J.W., et al., *Protein design – a hierarachical approach*. Science, 1995. **270**(5238): 935–941.
126. Kirshenbaum, K., R.N. Zuckermann, and K.A. Dill, *Designing polymers that mimic biomolecules*. Current Opinion in Structural Biology, 1999. **9**(4): 530–535.
127. Kim, S. and P.A. Coulombe, *Intermediate filament scaffolds fulfill mechanical, organizational, and signaling functions in the cytoplasm*. Genes & Development, 2007. **21**(13): 1581–1597.
128. Herrmann, H., et al., *Intermediate filaments: from cell architecture to nanomechanics*. Nature Reviews Molecular Cell Biology, 2007. **8**(7): 562–573.
129. Ackbarow, T., S. Keten, and M.J. Buehler, *A multi-timescale strength model of alpha-helical protein domains,* Journal of Physics: Condensed Matter, 2009. **21**: 035111.
130. Bell, G.I., *Models for specific adhesion of cells to cells*. Science, 1978. **200**(4342): 618–627.
131. Evans, E.A. and D.A. Calderwood, *Forces and bond dynamics in cell adhesion*. Science, 2007. **316**(5828): 1148–1153.
132. Evans, E., *Probing the relation between force – lifetime – and chemistry in single molecular bonds*. Annual Review of Biophysics and Biomolecular Structure, 2001. **30**: 105–128.
133. Evans, E.B., *Looking inside molecular bonds at biological interfaces with dynamic force spectroscopy*. Biophysical Chemistry, 1999. **82**(2–3): 83–97.
134. Merkel, R., et al., *Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy*. Nature (London), 1999. **379**(6714): 50–53.
135. Dudko, O.K., G. Hummer, and A. Szabo, *Intrinsic rates and activation free energies from single-molecule pulling experiments*. Physical Review Letters, 2006. **96**(10): 108101.
136. Makarov, D.E., *Unraveling individual molecules by mechanical forces: Theory meets experiment*. Biophysical Journal, 2007. **92**(12): 4135–4136.
137. West, D.K., P.D. Olmsted, and E. Paci, *Mechanical unfolding revisited through a simple but realistic model*. Journal of Chemical Physics, 2006. **124**(15): 154909.
138. Erdmann, T. and U.S. Schwarz, *Stability of adhesion clusters under constant force*. Physical Review Letters, 2004. **92**(10): 108102.
139. Lantz, M.A., et al., *Stretching the alpha-helix: A direct measure of the hydrogen-bond energy of a single-peptide molecule*. Chemical Physics Letters, 1999. **315**(1–2): 61–68.
140. Kageshima, M., et al., *Insight into conformational changes of a single alpha-helix peptide molecule through stiffness measurements*. Chemical Physics Letters, 2001. **343**(1–2): 77–82.
141. Dudko, O.K., et al., *Extracting kinetics from single-molecule force spectroscopy: Nanopore unzipping of DNA hairpins*. Biophysical Journal, 2007. **92**(12): 4188–4195.
142. Keten, S. and M.J. Buehler. *Strength limit of entropic elasticity in beta-sheet protein domains.* Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), 2008. **78**(6): 061913.
143. Griffith, A.A., *The phenomenon of rupture and flows in solids*. Philosophical Transactions of the Royal Society of London A, 1920. **221**: 163–198.
144. Sheu, S.-Y., et al., *Energetics of hydrogen bonds in peptides*. PNAS, 2003. **100**(22): 12683–12687.
145. Rief, M., et al., *Single molecule force spectroscopy of spectrin repeats: Low unfolding forces in helix bundles*. Journal of Molecular Biology, 1999. **286**(2): 553–561.
146. Law, R., et al., *Influence of lateral association on forced unfolding of antiparallel spectrin heterodimers*. Journal of Biological Chemistry, 2004. **279**(16): 16410–16416.
147. Lenne, P.F., et al., *Stales and transitions during forced unfolding of a single spectrin repeat*. FEBS Letters, 2000. **476**(3): 124–128.
148. Law, R., et al., *Cooperativity in forced unfolding of tandem spectrin repeats*. Biophysical Journal, 2003. **84**(1): 533–544.

149. Law, R., et al., *Pathway shifts and thermal softening in temperature-coupled forced unfolding of spectrin domains*. Biophysical Journal, 2003. **85**(5): 3286–3293.
150. Bernstein, F.C., et al., *Protein data bank – computer-based archival file for macromolecular structures*. Journal of Molecular Biology, 1977. **112**(3): 535–542.
151. Kolano, C., et al., *Watching hydrogen-bond dynamics in a beta-turn by transient two-dimensional infrared spectroscopy*. Nature, 2006. **444**(7118): 469–472.
152. Grandbois, M., et al., *How strong is a covalent bond?* Science, 1999. **283**(5408): 1727–1730.
153. Bustamante, C., et al., *Entropic elasticity of lambda-phage DNA*. Science, 1994. **265**(5178): 1599–1600.
154. Marko, J.F. and E.D. Siggia, *Stretching DNA*. Macromolecules, 1995. **28**(26): 8759–8770.
155. Zhuang, X., *Molecular biology: Unraveling DNA condensation with optical tweezers*. Science, 2004. **305**(5681): 188–190.
156. Lang, M., *Lighting up the mechanome*, in *Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2007 Symposium*. 2008, National Academy of Engineering of the National Academies.
157. Ebenstein, D.M. and L.A. Pruitt, *Nanoindentation of biological materials*. Nano Today, 2006. **1**(3): 26–33.
158. Bozec, L., et al., *Atomic force microscopy of collagen structure in bone and dentine revealed by osteoclastic resorption*. Ultramicroscopy, 2005. **105**(1–4): 79–89.
159. Guzman, C., et al., *Exploring the mechanical properties of single vimentin intermediate filaments by atomic force microscopy*. Journal of Molecular Biology, 2006. **360**(3): 623–630.
160. Yuan, C.B., et al., *Energy landscape of streptavidin-biotin complexes measured by atomic force microscopy*. Biochemistry, 2000. **39**(33): 10219–10223.
161. Sun, Y.L., Z.P. Luo, and K.N. An, *Stretching short biopolymers using optical tweezers*. Biochemical and Biophysical Research Communications, 2001. **286**(4): 826–830.
162. Sazonova, V., et al., *A tunable carbon nanotube electromechanical oscillator*. Nature, 2004. **431**(7006): 284–287.
163. Thorsen, T., S.J. Maerkl, and S.R. Quake, *Microfluidic large-scale integration*. Science, 2002. **298**(5593): 580–584.
164. Whitesides, G.M. and B. Grzybowski, *Self-assembly at all scales*. Science, 2002. **295**(5564): 2418–2421.
165. Yan, H., et al., *DNA-templated self-assembly of protein arrays and highly conductive nanowires*. Science, 2003. **301**(5641): 1882–1884.
166. Rho, J.Y., L. Kuhn-Spearing, and P. Zioupos, *Mechanical properties and the hierarchical structure of bone*. Medical Engineering and Physics, 1998. **20**(2): 92–102.
167. Currey, J.D., *Mechanical properties of mother of pearl in tension*. Proceedings of the Royal Society of London. Series B, Biological Sciences, 1977. **196**(1125): 443–463.
168. Menig, R., et al., *Quasi-static and dynamic mechanical response of Strombus gigas (conch) shells*. Materials Science & Engineering A, 2001. **297**(1–2): 203–211.
169. Tesch, W., et al., *Graded microstructure and mechanical properties of human crown dentin*. Calcified Tissue International, 2001. **69**(3): 147–157.
170. Ritchie, R., M.J. Buehler, P. Hansma. *The strength and toughness of bone,* Physics Today, 2009. **62**(6): 41-47.
171. Taylor, D., J.G. Hazenberg, and T.C. Lee, *Living with cracks: Damage and repair in human bone*. Nature Materials, 2007. **6**(4): 263–266.
172. Nalla, R.K., J.J. Kruzic, and R.O. Ritchie, *On the origin of the toughness of mineralized tissue: Microcracking or crack bridging?* Bone, 2004. **34**(5): 790–798.

173. Gao, H., et al., *From the cover: Materials become insensitive to flaws at nanoscale: Lessons from nature*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(10): 5597.
174. Mann, S., et al., *Crystallization at Inorganic-organic Interfaces: Biominerals and biomimetic synthesis*. Science, 1993. **261**(5126): 1286–1292.
175. Gilbert, P., M. Abrecht, and B.H. Frazer, *The organic-mineral interface in biominerals*. Reviews in Mineralogy and Geochemistry, 2005. **59**(1): 157–185.
176. Smith, B.L., et al., *Molecular mechanistic origin of the toughness of natural adhesives, fibres and composites*. Nature, 1999. **399**(6738): 761–763.
177. Broedling, N.C., et al., *The strength limit in a bio-inspired metallic nanocomposite*. Journal of Mechanics and Physics of Solids, 2008. **56**(3): 1086–1104.
178. Sen, D. and M.J. Buehler, *Crystal size controlled deformation mechanism: Breakdown of dislocation mediated plasticity in single nanocrystals under geometric confinement*. Physical Review B, 2008. **77**(19): 195439.
179. Sen, D. and M.J. Buehler, *Shock loading of bone-inspired metallic nanocomposites*. Solid State Phenomena, 2008. **139**: 11–22.
180. Whitesides, G.M., J.P. Mathias, and C.T. Seto, *Molecular self-assembly and nanochemistry: A chemical strategy for the synthesis of nanostructures*. Science, 1991. **254**(5036): 1312–1319.
181. Whitesides, G.M. and B. Grzybowski, *Self-assembly at all scales*. Science, 2002. **295**: 2418–2421.
182. Bejan, A., *Constructal theory: From thermodynamic and geometric optimization to predicting shape in nature*. Energy Conversion and Management, 1998. **39**(16–18): 1705–1718.
183. Kim, P., et al., *Thermal Transport Measurements of Individual Multiwalled Nanotubes*. Physical Review Letters, 2001. **87**(21): 215502.
184. Balandin, A.A., et al., *Superior thermal conductivity of single-layer graphene*. Nano Letters, 2008. **8**(3): 902–907.
185. Meng, G., et al., *Controlled fabrication of hierarchically branched nanopores, nanotubes, and nanowires*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(20): 7074–7078.
186. Shinde, S.L. and J.S. Goela, eds. *High Thermal Conductivity Materials*. 2004, New York: Springer, p. 271.
187. Langer, R. and D.A. Tirrell, *Designing materials for biology and medicine*. Nature, 2004. **428**(6982): 487–492.
188. Zhao, X.J. and S.G. Zhang, *Designer self-assembling peptide materials*. Macromolecular Bioscience, 2007. **7**(1): 13–22.
189. Holland, J.H., *Hidden Order – How Adaptation Builds Complexity*. 1995, Reading, MA: Helix Books.
190. Ackbarow, T. and M.J. Buehler, *Hierarchical coexistence of universality and diversity controls robustness and multi-functionality in protein materials.* Journal of Computational and Theoretical Nanoscience, 2008. **5**(7): 1193–1204.
191. Cui, X.Q., et al., *Biocatalytic generation of ppy-enzyme-CNT nanocomposite: From network assembly to film growth*. Journal of Physical Chemistry C, 2007. **111**(5): 2025–2031.
192. Hule, R., D. Pochan, *Polymer nanocomposites for biomedical application*. MRS Bulletin, 2007. **32**(4): 5.
193. Winey, K.I., Vaia R.A., *Polymer nanocomposites*. MRS Bulletin, 2007. **32**(4): 5.
194. Petka, W.A., et al., *Reversible hydrogels from self-assembling artificial proteins*. Science, 1998. **281**(5375): 389–392.
195. Smeenk, J.M., et al., *Controlled assembly of macromolecular beta-sheet fibrils*. Angewandte Chemie-International Edition, 2005. **44**(13): 1968–1971.

196. Zhao, X.J. and S.G. Zhang, *Molecular designer self-assembling peptides*. Chemical Society Reviews, 2006. **35**(11): 1105–1110.
197. Mershin, A., et al., *A classic assembly of nanobiomaterials*. Nature Biotechnology, 2005. **23**(11): 1379–1380.
198. Aebi, U., et al., *The nuclear lamina is a meshwork of intermediate-type filaments*. Nature, 1986. **323**(6088): 560–564.
199. Buehler, M.J., *Hierarchical chemo-nanomechanics of proteins: Entropic elasticity, protein unfolding and molecular fracture*. Journal of Mechanics of Materials and Structures, 2007. **2**(6): 1019–1057.
200. Ashby, M.F., et al., *The mechanical properties of natural materials. I. Material property charts*. Proceedings of the Mathematical and Physical Sciences, 1995. **450**(1938): 123–140.
201. Fratzl, P., et al., *Structure and mechanical quality of the collagen–mineral nano-composite in bone*. Journal of Materials Chemistry, 2004. **14**(14): 2115–2123.

CHAPTER 18

COMPUTATIONAL MOLECULAR BIOMECHANICS: A HIERARCHICAL MULTISCALE FRAMEWORK WITH APPLICATIONS TO GATING OF MECHANOSENSITIVE CHANNELS OF LARGE CONDUCTANCE

XI CHEN[1], AND QIANG CUI[2]

[1]*Department of Earth and Environmental Engineering, Columbia University, New York, NY 10027, USA, e-mail: xichen@columbia.edu*

[2]*Department of Chemistry, University of Wisconsin, Madison, WI 53706, USA, e-mail: cui@chem.wisc.edu*

Abstract: Understanding the mechanism of mechanobiological processes at the molecular level is an important challenge in modern biophysics. Despite recent advances in experimental and numerical techniques, the intrinsic multiscale nature of mechanobiological processes makes it difficult to meet such challenge in many systems of interest. Recently, a continuum-mechanics based hierarchical modeling and simulation framework has been established and applied to study the mechanical responses and gating behaviors of a prototypical system, the mechanosensitive channel of large conductance (MscL) in bacteria *Escherichia coli* (*E. coli*), from which several putative gating mechanisms have been testified and new insights deduced. This article reviews these latest findings and suggests possible improvements for future modeling work. The computationally efficient and versatile continuum-based protocol is expected to make contributions to a variety of mechanobiology problems

Keywords: Molecular biomechanics, Computation, Mechanosensitive channel

18.1. INTRODUCTION

Many challenging problems in molecular biomechanics require the development of new efficient computational tools. For example, cellular mechanotransduction, which is the mechanism by which cells convert mechanical stimuli into biochemical responses, has been a research focus for several decades owing to its significance in many physiological functions in living organisms, such as touching, balance, and hearing, among others [1–4]. Despite its significance, the intrinsic mechanisms by which the cell responses to external stimuli are not yet well understood. During mechanotransduction, the external force at macroscopic scale can transcend down

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 535–556.
DOI 10.1007/978-1-4020-9785-0_18,

multiple length scales to the microscopic mechanical behaviors of biomolecules and their assemblies. Therefore, understanding the molecular basis for mechanotransduction requires bridging multiple length- and time-scales along with an efficient treatment of the complex structure of biomolecules; this is still challenging for experimental studies, and thus important insights are sought from modeling and simulations. The present numerical schemes based on all-atom simulations are limited to simple loading modes occurring at nanometer and nanosecond scales. An efficient multiscale framework that can both capture sufficient molecular details and deal with complex loadings over multiple scales is undoubtedly a valuable supplement to the conventional experimental and modeling methods.

Recently, a top-down computational framework has been developed and referred to as the molecular dynamics-decorated finite element method (MDeFEM) [5–7]. In essence, MDeFEM effectively models the biomolecule and its assemblies as integrated structures; the mechanical properties of each structural component as well as their interactions are derived from atomistic simulations, and the entire structure is then solved using FEM. This multiscale approach can effectively bridge the gap of previous numerical techniques, allowing important mechanistic insights of mechanobiology to be deduced.

In this chapter, based on the model system of mechanosensitive channel of large conductance (MscL) in *E. coli*, the fundamental gating mechanisms are deduced using the continuum-based MDeFEM framework [5–7]. The results are compared with pervious studies from experimental, theoretical and numerical perspectives, and the obtained insights are potentially relevant to understanding a host of other biomolecular systems, including the mechanosensitive (MS) channels in higher organisms. Further improvements of the present multiscale model are also proposed and it is envisioned that such an improved framework will find great value in computational molecular biomechanics.

18.2. BRIEF OVERVIEW OF MECHANOSENSITIVE (MS) CHANNELS

18.2.1. Structural Components of MS Channel of Large Conductance (MscL)

Although some long-term mechanotransduction events such as tissue remodeling involve the consequence of altered gene expression, most cellular responses to mechanical forces are due to the MS channels [1, 4, 8]. In bacteria, MS channels respond to load perturbation applied to the cell membrane or other membrane-associated components and act as “safety” valves and facilitate the permeation of small ions and water molecules [4]. One of the most studied MS channels up to date is bacterial MscL, for which the structure has been characterized at atomic scale by x-ray crystallography [9] or homology modeling [10]. In Figure18-1a and b (top/side views), the structure of *E. coli*-MscL in the closed state is shown, which was developed by homology modeling [10, 11] based on the *X*-ray crystal

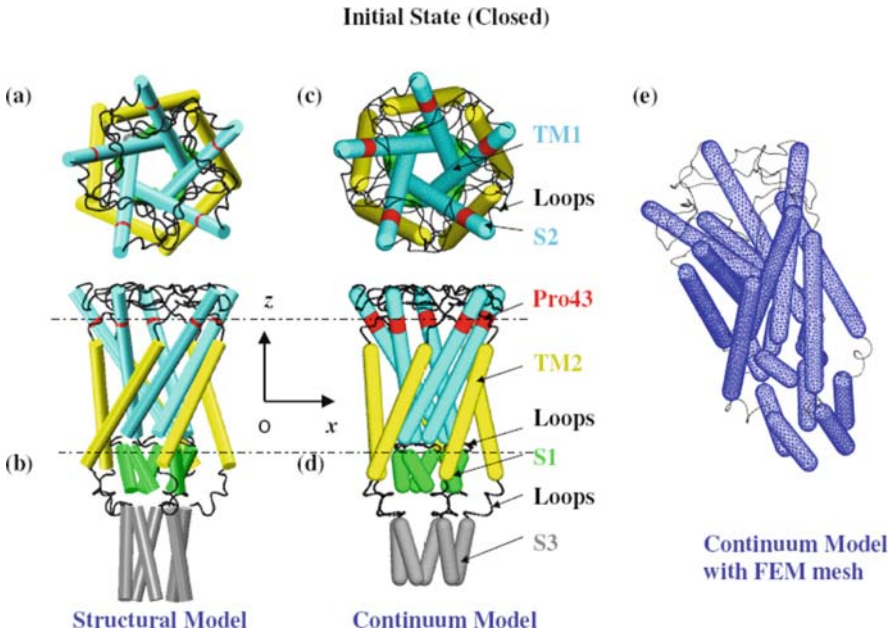


Figure 18-1.E. coli-MscL: (**a**) *top view* and (**b**) *side view* of the closed structure of the homology model (**c**) *top view* and (**d**) *side view* of the MDeFEM model; (**e**) mesh of the protein. The *dash line* indicates the initial location of lipid membrane. The protein components include: the transmembrane TM1 bundle (*sky blue*) and TM2 subunits (*yellow*), cytoplasmic S1 helices (*green*) and S3 helices (*grey*), and they are connected by periplasmic and cytoplasmic loops (black)

structure of MscL in bacteria *Mycobacterium Tuberculosis* (*Tb*) [9] and other available experimental data [12]. The structure of *E. coli*-MscL is of five-fold symmetry, and the residues on top of the transmembrane helices are connected by periplasmic loops (black), whereas those at the bottom of the transmembrane helices are linked to cytoplasmic helices via cytoplasmic loops (black). Among the transmembrane helices that directly interact with the membrane, the TM1 bundle (sky blue) consists of five longer subunits that form an inner gate, and the five TM2 subunits (yellow) form the outer bundle. The dash line in Figure 18-1b indicates the initial location of lipid membrane. In *E. coli*-MscL, TM1 and TM2 helices correspond to residues Asn 15-Gly 50 and Val 77-Glu 107, respectively. There is a break in TM1 due to Pro43 (red) near the top of the TM1 helix; in the literature, the segment above Pro43 is sometimes referred to as the S2 helices [10]. The cytoplasmic domain is composed of gates formed by S1 helices (green) and S3 helices (grey), which correspond to residues Ile 3-Met 12 and Lys 117-Arg 135, respectively. Among the three inner valves formed by the TM1, S1 and S3 helix bundle, the transmembrane pore enclosed by TM1 helices is the most important, which determines the ion flux that passes through and can be estimated by measuring the electric current experimentally [12]. Hence, an effective radius of the channel could becalculated from the area

of the pentagon projection on the membrane plane formed by the principle axes of the TM1 helix bundle, which is ~6.5 Å for the *E. coli*-MscL in its closed state.

The gating transitions of MS channels can be stimulated by external force through the lipid environments. Membrane-activated gating behaviors were first found by Kung's group by carrying out patch-clamp experiment on lipid vesicles [13]; since cytoskeleton and other membrane proteins were removed, the mechanical deformation of lipid was demonstrated as the key factor for MS channel gating. The crucial role of membrane was also validated using a thermodynamics analysis [14], where the free-energy of the lipid bilayer deformation was shown to be on the same order as the energy barrier required for gating. Other experiments [15, 16] and numerical studies [17, 18] showed that MS channels are sensitive to lipid composition. Therefore, a study that explicitly considers lipid membrane is required for the mechanistic analysis of gating of MS channels.

A lipid bilayer membrane is composed of phospholipids and it forms a natural barrier between the inside and outside of the cell [19]. The exposed head groups of phospholipids are hydrophilic, and the tails are hydrophobic toward the center of lipid bilayer. An example of dilauroyl-phosphatidylethanolamine (DLPE) lipid is shown in Figure 18-2a, which bounds an *E. coli*-MscL. Under the physiological condition, the lipid bilayer is fluidic-like: it is incapable of bearing shear stress and cannot sustain large strains, and several percent of area expansion may rupture

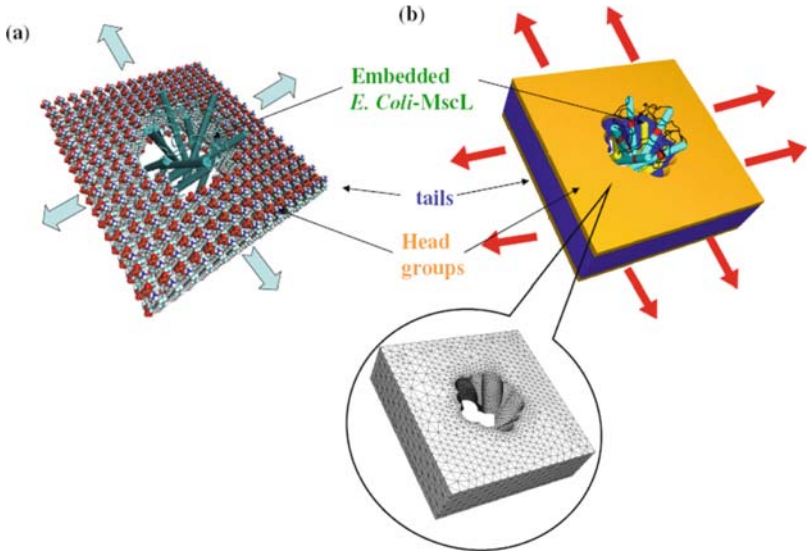


Figure 18-2. The assembled protein/lipid system: (**a**) the "cartoon" representation of *E. coli*-MscL and all-atom representation of a lipid, with a schematic of equi-biaxial loading applied at the lipid boundary. (**b**) continuum-mechanics based model of *E. coli*-MscL-lipid system. The finite element mesh of the trilayer lipid is shown in the zoom-in view. The integrated protein structure (established from the continuum approach) is embedded in a continuum lipid slab (based on a sandwich model). The effective mechanical properties of the continuum components as well as their interactions are derived from MD simulations

the membrane [1]. Nevertheless, under the assumption that it behaves like an elastic solid, the effective mechanical properties of lipid (upon tension, bending, etc.) can be estimated through relevant loads and deformation perturbations, and used for modeling purposes [14]. For instance, the area expansion coefficient of a lipid bilayer is the membrane tension per relative change of its unit area and is related with an effective Young's modulus of the membrane. Lipid membrane is nearly incompressible [20], which implies that its effective Poisson's ratio is close to 0.5. Besides in-plane tension, bending of lipid bilayer affects its curvature, and such contribution from an effective bending stiffness has been suggested to be important for the gating of MS channels [21]. In addition, the lipid property may be inhomogeneous across the thickness, since distinct peaks of lateral pressure are found in the head group and tail regions [18], and modification the pressure profile can lead to different channel gating characteristics [22–25].

Under the physiological condition, diverse external stimuli [1] may be exerted on a cell including but not limited to steady-state contacts, high-frequency vibrations, fluid shear stresses, etc.; these external stimuli may be superimposed with those generated internally, such as cytoskeletal polymerization, osmotic and hemodynamic pressure. It is interesting to investigate which stimulus is more relevant to mechanotransduction pathways. Since the forces acting on a MS channel are transferred through the lipid, the effect of various membrane deformation modes (e.g., in-plane tension, bending, etc.) vs. gating transitions may be explored. The schematic in Figure 18-2a shows the example of equi-biaxial tension, which is the most studied membrane deformation mode that can be triggered by variation of the osmotic pressure.

18.2.2. Previous Experimental and Theoretical Investigations

Pressure-activated, voltage dependent and ion selective properties of *E. coli*-MscL have been extensively studied [26], as well as their gating behaviors using electrophysiological characterization [13]. The relationship between channel opening probabilities and membrane tension was reported in [27] and a five-subconductance-states model was established, which showed that the tension-dependent conformational transition is primarily attributed to the pore area variation that occurred between the closed state and the first subconductance state [12]. Based on experimental constraints and known structural features of channel proteins, structural models for the gating transition of *Tb*-MscL and *E. coli*-MscL upon equi-biaxial tension have been established [28]. These models include 13 conformational states ranging from the fully closed state (when the effective pore radius a is about 6.5 Å) to an open conformation (where the maximum conductance can be measured experimentally) with a=19 Å; these models are consistent with results from cysteine cross-linking experiments [10].

Theoretical studies are useful for postulating the principles behind the gating transition of MscL. Philips et al. established a lipid-centric model [14] where the system is dominated by hydrophobic mismatch and lipid tension; the protein deformation was completely ignored. This model was improved by incorporating other

triggers, such as the change of membrane curvature and mid-plane deformation between the closed and opened states [24]; more recently, it was further extended to study the issue of gating co-operativity [29]. Although these thermodynamics-based models provided useful insights (such as the importance of lipid deformation) into the common features of MS channels, the lack of sufficient structural details makes it difficult to evaluate their validity for a specific system (which is particularly important in biological systems where certain atomistic features are crucial to function). For example, the proteins were treated as either rigid or overly simplified objects [14, 24, 29], thus their deformation energy contributions were largely ignored. Moreover, key parameters in these models were usually not derived from detailed simulations or experiments.

18.2.3. Previous Numerical Approaches

Effective numerical simulations are a powerful alternative for exploring the fundamental principles of molecular biomechanics. Comparing to analytical modeling, numerical simulation can incorporate sufficient structural details to propose or verify various theoretical hypotheses as well as to improve the model. In addition, numerical experiments can be manipulated in a precise way, which may help interpret experimental data and stimulate new experiments.

All-atom simulations based on molecular dynamics (MD) are useful to explore the gating transitions at the finest scale [30]. However, they are prohibitively expensive especially when the protein, lipid membrane, and surrounding solvent are explicitly considered [31]. MD simulation was first applied to study the gating of *Tb*-MscL [32]; limited by the accessible time-scale (3 ns), the pore size was barely changed. To accelerate structural transitions, the protein deformation was decoupled from the membrane, and external forces were applied directly to the protein so as to assist gating. Steered molecular dynamics (SMD) was employed to study *E. coli*-MscL upon equi-biaxial tension [30], where the steering forces, estimated from the lateral and normal pressure profiles exerted by the deformed bilayer, were added to selected boundary atoms of the protein. Despite such bias, the channel merely opened to a=9.4 Å after 12 ns of simulation, which highlights the limitation of atomistic MD simulations in the context of probing the channel gating process.

Another artificially accelerated approach was based on the targeted MD (TMD) [33]. The lipid bilayer membrane was completely ignored and a holonomic constraint was used to drive the opening of an *E. coli*-MscL. The constraining force, however, can be unrealistically large [34] which makes the targeted MD useful only in a qualitative sense [33]. A similar approach was carried out by Bilston and coworkers [35] on a *Tb*-MscL without explicit lipid membrane, and the opening of MscL was negligible after 150 ps of simulation.

During a simulation with short physical time, the applied external forces can lead to unrealistic protein motions [16]. To circumvent such problem, Meyer et al. [36] studied the conformational transitions of *E. coli*-MscL in a pre-curved lipid membrane. Interesting observations were made through the 9.5 ns simulation, where the major structural rearrangements occurred in the periplasmic loops and extracelluar

helices. Although the channel radius increment was still limited, this technique has successfully combined bending and equi-biaxial tension modes, and the transition of *E. coli*-MscL occurred locally in absence of global external forces.

Despite various improvements, the all-atom-based simulations are still computationally intensive and the short simulation time may be inadequate for statistical sampling. Hence, to overcome the shortcomings of the all-atom simulations, multiscale approach must be adopted. The concurrent simulations developed for crystals [37], however, encounters difficulties on biomaterials especially for the coarse grain modeling of soft tissues. A coarse grained model [38] was developed recently based on a thermodynamics parameterization and the gating of *Tb*-MscL was simulated; however, further validation of the coarse-grained force field is needed. The computational cost of the particle-based multiscale model is still high, and these types of models are difficult to employ for studying deformations involving large length scales or complex loading modes.

The limitations of present atomistic and coarse-grained simulations motivated us to develop a novel continuum mechanics-based modeling and simulation framework for mechanobiology [5], which can not only bridge multiple length scales and adapt complex loadings, but also include sufficient molecular details to capture some of the most important characteristics of the specific system. Being top-down in nature, the molecular dynamics-decorated finite element method (MDeFEM) complements the traditional bottom-up all-atom/coarse-grained simulations. In the hierarchical approach, the biomolecule and its assemblies are modeled as integrated continuum structures incorporating some of the most important structural details (and redundant atomic details are ignored). This is motivated by the fact that the mechanical deformation of a biomolecular system is likely dictated by the superposition of low-energy modes, which can be well described by the "collective" behavior of its structural motifs (via phenomenological mechanical properties) and the local chemical/atomistic details are likely less important. Thus, the MDeFEM can efficiently treat large deformations at length scales and complex deformation modes not accessible to conventional MD simulations, while still retaining some key features from the atomic scale; the method can also be improved by choosing different levels of refinement (or decoration) based on atomistic simulations.

In this chapter, we discuss the gating mechanism of *E. coli*-MscL based on numerical simulations with MDeFEM (Figure 18-2b) [5, 7]; more details of modeling and computational methods are described below. The underlying gating mechanisms of *E. coli*-MscL [7] may provide useful insights for the function of other mechanosensitive channels and cellular mechanotransduction.

18.3. CONTINUUM-BASED APPROACH: MODEL AND METHODS FOR STUDYING MSCL

In Figure 18-1c and d, the top/side views of the continuum model of *E. coli*-MscL are shown [5]; the geometries of all continuum components are measured from the closed structure of the homology model [11] (Figure 18-1a and b). Each helix (TM1/TM2/S1/S2/S3) is modeled as a three-dimensional elastic cylinder with

diameter 5 Å (a typical value for the main chain of an α-helix). As a first order approximation, the helix is taken to be homogeneous and isotropic, and the elastic properties remain constant during the gating transition. Due to its possible important contribution, Pro43 [10] is treated explicitly in the helix model. The materials properties of the helices are calibrated by matching results of normal mode analysis (NMA) at the atomistic and continuum levels. The atomistic NMA is carried out using the CHARMM19 force field [39, 40], which is more appropriate here because the calculations are done in vacuum; the effect of solvation is approximated by adopting a distance-dependent dielectric constant in electrostatic calculations, and no solvent damping effect is considered to be consistent with the continuum calculations. The Young's modulus is varied such that the eigenvalues and eigenvectors for the three lowest-frequency modes computed at the continuum level best fit the results from the atomistic normal mode calculation [5]. In general, the Young's moduli of helices are within the range found in previous simulation studies [41].

The loops are taken to be quasi-one dimensional elastic springs, whose mechanical properties are also assumed to residue-independent. Their mechanical properties are also assumed to be homogeneous and obtained by the similar normal mode fitting at the atomistic and continuum levels as discussed above for the helices [5]. The spring constants for the loops are rather stiff, which is also consistent with previous studies [30] and supports their possible importance during gating.

A lipid bilayer can be effectively modeled as a sandwich plate structure (Figure 18-2b) by considering the different roles played by the head and tail regions in transducing mechanical stress [17, 18]. This is motivated by the natural difference between the chemical and physical properties in these regions; e.g., it has been well established that the lateral pressure profile of lipid bilayers has distinct peaks at the interface (neck) between the head and tail regions, and that modifying the pressure profile can lead to different gating characteristics for the MS channels [18]. Each layer (head group or tail) is assumed to be homogeneous and elastic, whose effective thickness and elastic constants are fitted based on published MD results [18] by letting the strain energies to be equivalent during in-plane expansion. To host the MS channel, a cavity with conforming shape is created in the membrane. Based on such a continuum model, the assembled continuum structure of an *E. coli*-MscL inside a lipid bilayer is shown in Figure 18-2b.

With a continuum-mechanics based representation, the interactions among atoms within each continuum component are not computed explicitly because the corresponding energy is implicitly represented via the phenomenological mechanical properties described in the last subsection; this is one reason that the computational cost associated with the continuum framework is substantially lower than all-atom simulations.

In MDeFEM, the structural components of biomolecules are integrated together through non-bonded interactions (electrostatic and van der Waals). For different pairs of interactions (e.g. between helices and those between helix and lipid), pairwise Lennard-Jones potentials are employed to describe the non-bonded forces [5], which can be calibrated by calculating and matching potential energies at atomistic

and continuum levels. Specifically, the non-bonded interactions between helices and that between helix and lipid can be represented by,

$$E_{\text{int}}(\alpha) = C\left[\frac{n}{m}\left(\frac{d_0}{\alpha}\right)^m - \left(\frac{d_0}{\alpha}\right)^n\right],$$

where $E_{\text{int}}(\alpha)$ is the non-bonded interaction energy between the surfaces of two continuum components, which include contributions from both electrostatic forces and van der Waals interactions. For any given pair of interaction, d_0 is the (shortest) initial equilibrium distance between the two surfaces, and α is the surface distance between two deformed surfaces. Both m and n are positive integers that account for repulsive and attractive terms, respectively, with $n < m$ in general. The value of d_0 depends on the different types of continuum components involved. The parameters including the "well-depth", C, and the exponents (n, m), are calculated based on fitting to energy calculations using atomistic molecular mechanics force field [5]. For each pair of helices, the interaction energy in the vacuum is calculated using the polar-hydrogen set of CHARMM19 force field. Calculations are done for different combinations of helical pairs, which have effectively sampled many relative orientations. To estimate the helix-lipid interactions, the insertion energy profiles are computed when a single helix (TM1 or TM2) is gradually transferred in and out of an implicit membrane with varying orientations; an implicit dielectric model is used for the membrane to avoid the need of sampling a large set of lipid configurations.

Parameterization of helices, loops, lipid, and interactions is described in Ref. [5]. The integrated system is then meshed with finite elements (see example in Figure 18-1e for the mesh of protein bundle and that in Figure 18-2b for the elements near lipid cavity). Commercial software ABAQUS [42] is used for FEM analyses; the typical simulation time is only a few hours on a regular workstation with a single CPU.

18.4. GATING MECHANISMS OF MSCL AND INSIGHTS FOR MECHANOTRANSDUCTION

18.4.1. Effect of Different Loading Modes

In previous experimental and theoretical studies [16, 21, 24, 25, 27, 30], several loading modes have been postulated to be the triggers of MscL gating, including dilatational gating (equi-biaxial tension) and gating-by-tilting (axisymmetric bending). These potential gating mechanisms are examined with MDeFEM simulations [7] to explore mechanotransduction pathways under different deformation modes.

18.4.1.1. Gating Behaviors Upon Equi-Biaxial Tension

An equi-biaxial strain up to 21% is applied as a displacement boundary condition on the membrane [7]. Such a large strain is a result of modeling the lipid as a solid slab; despite such bias, an appropriate expansion of lipid cavity can be achieved, which

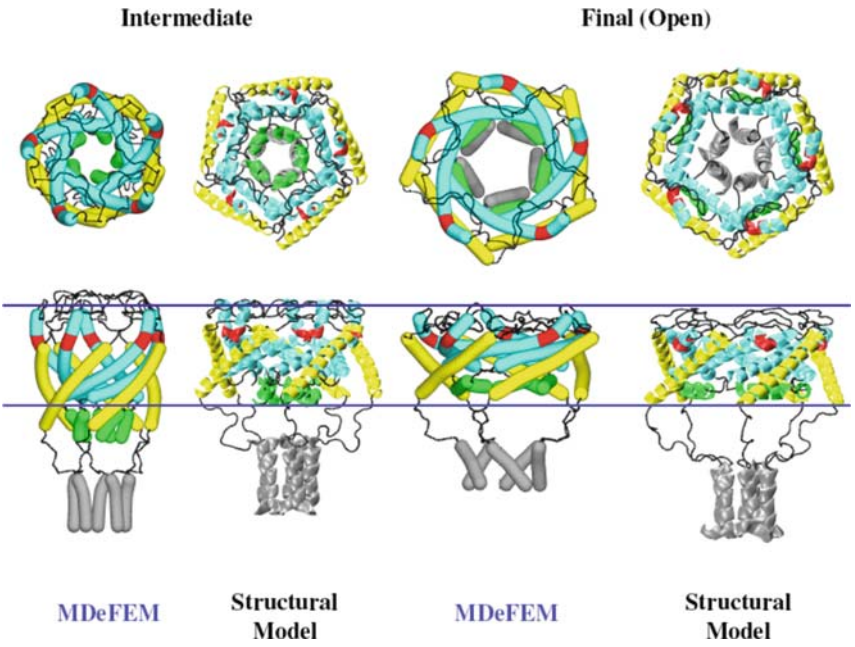


Figure 18-3. Snapshots of gating pathways of a single *E. coli*-MscL at half- and fully-opened states upon equi-biaxial tension: comparisons between the structural model and MDeFEM simulation (with maximum membrane strain 21%). Gating is primarily realized through the interaction between transmembrane helices and lipid, where the pore enclosed by the light blue TM1 helices is pulled open. Other cytoplasmic helices and loops follow the trajectories of the transmembrane helices

is necessary to accommodate the fully opened channel and enables us to explore the important aspects of the gating process that concern protein conformations.

The snap shots of the morphological transition of *E. coli*-MscL at intermediate (half-opened) and opened states [7] are given in Figure 18-3; they compare favorable to the structural model of Guy et al. [11], which reflects the effectiveness of the MDeFEM approach. It is not surprising that with the increase of membrane strain, the lipid cavity expands and the forces are transmitted to the transmembrane helices of the protein structure via non-bonded interactions. Consequently, the transmembrane region suffers most conformational changes characterized by their radial expansions, and the pore enclosed by the TM1 bundle opens up.

Besides the lateral expansion, visible shrinking in the thickness direction is also observed for the transmembrane region, which is related with significant tilting of the helices. The longer and more flexible TM1 bends more than TM2 helices; such significant deformation, which is required to maintain mechanical equilibrium during the gating process [7], remains to be verified by experimental studies with sufficient resolution (although the structural model [11] has predicted similar helix bending curvature, Figure 18-3).

The MscL pore is defined by the TM1 helices; thus the loops connecting TM1 and TM2 helices also impose constraints on the size of the pore. In addition, the S1

helices also affect the pore conformation via interaction with TM1 helices; during gating, the S1 bundle expands in the radial direction and they are lifted up towards the transmembrane region, which confirms the "swing-like" motions of N-terminus [43]. The expansion of the S1 helices, however, is smaller than that in the structural model; this might have been caused by the neglect of solvation contributions in the current model (see Section 18.6). Being far away from the transmembrane helices, the S3 assembly remains essentially unchanged; this finding is in agreement with the later version of the structural model [11] (as opposed to the previous version [28]), suggesting that the S3 helices may be less important in terms of their mechanics roles (although they may bear other biochemical functions). The present system is resilient and can recover its closed state when the membrane stress is removed.

In Figure 18-4a, the percentage increment of the effective pore radius of *E. coli*-MscL, *a*, is calculated from MDeFEM simulation as a function of membrane strain (the open-square curve) [7]. According to elasticity theory, the relationship between membrane strain and lipid cavity expansion is linear [5], and that results in the monotonic behavior of pore radius increment; the minor iris-like features in Figure 18-4a are due to the many-body interactions that affect equilibrium.

At small strain, the variation of channel radius agrees well between MDeFEM [7] and SMD simulations (the open-circle curve in Figure 18-4a) [30]. This demonstrates that at least qualitatively, the continuum-based approach model has a reasonable description for the forces involved in the gating process, and it can reach the fully opened state with a fraction of computational cost (several hours on a PC workstation vs. months on a supercomputer). Thus, MDeFEM holds promising advantages over all-atom simulations in terms of the accessible length- and time-scales.

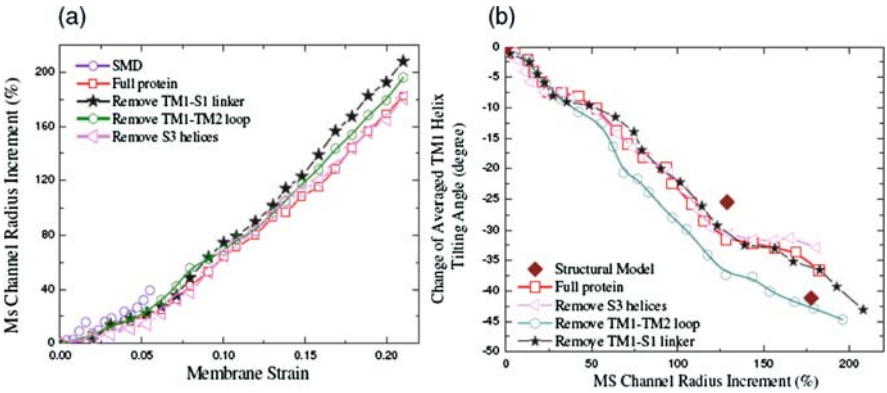


Figure 18-4. Equi-biaxial tension: (**a**) the evolution of the effective pore radius (enclosed by TM1 helices) of *E. coli*-MscL versus membrane strain (**b**) the change of TM1 helix tilting angle as a function of the effective MS channel radius. The results obtained from MDeFEM simulation are compared with those from steered MD simulation (at small strains) and the structural model. In addition, various structural motifs are removed to explore the effect of these protein components during gating

Another parameter useful for characterizing the pore shape variation is the averaged tilting angle of TM1 helices (with respect to the membrane plane), which decreases monotonically with pore radius (open-square curve in Figure 18-4b) [7]. Again, the MDeFEM results agree qualitatively with that of the structural model in the intermediate and opened states (solid-diamond symbols) [11]. It is worth noting that the MDeFEM model is based mostly on the closed structure of MscL (unlike that in many other simulations where the final state must also be given to explore the pathways in between [33]), and thus such an agreement is quite remarkable.

Although the MDeFEM study has demonstrated some promising results and agreements with the structural model, the largely monotonic behaviors of pore radius and helix tilting angle in Figure 18-4 illustrate a main limitation of the current model, for which the effective energy surface is essentially downhill toward the open state in the presence of external load. This is inconsistent with the free energy profile estimated in Ref. [12], which involves various intermediate states separated by notable energy barriers; moreover, at the opened state, the channel radius is essentially insensitive to a wide range of tensions in experiments. We believe that further refinements, such as incorporating the effect of solvation forces [38, 44, 45], may help to make the MDeFEM approach more realistic (see Section 18.6).

18.4.1.2. Gating Behaviors Upon Bending

Membrane bending is a commonly encountered deformation mode in a flexible cellular structure and becomes prominent during cell adhesion/contact. In order to study the pure bending behavior (i.e., decoupled with membrane stretching), a four-point bend flexure of a circular membrane is studied [7]. With respect to the insert in Figure 18-5a, one can define the cone angle, β, with reference to the effective radii of the five TM1 helices at the locations that correspond to the surfaces of the lipid membrane. When the membrane is bent upwards, the cone angle decreases almost monotonically with bending moment, and the TM1 helices become more upright at the final bending stage (Figure 18-5a).

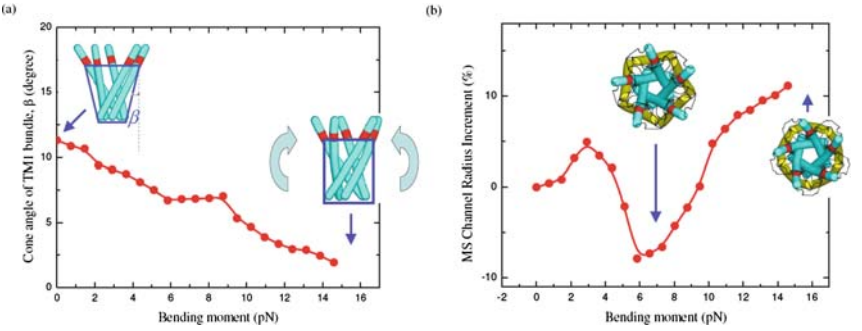


Figure 18-5. Axisymmetric pure bending as the membrane is bent upwards: (**a**) change of the TM1 helix bundle cone angle β as a function of the line bending moment, with the inserts showing configurations of the TM1 bundle at different instants (**b**) increment of the effective pore radius versus the line bending moment, with the inserts showing configurations (*top view*) of the MscL at different instants

Whether pure bending is an effective mode to promote gating can be explored from the effective pore radius evolution with bending moment as shown in Figure 18-5b. Despite the rotation of the wall of the lipid cavity, the averaged cavity radius throughout thickness remains about the same during pure bending; therefore, the overall variation of the channel radius is small. The snapshots of the channel at half and maximum bending moment are also given in Figure 18-5b, which show that despite the protein conformational change, the transmembrane pore radius is only moderately enlarged. The evolution of pore radius shows a zigzag pathway with bending moment, which is attributed to the many-body interactions among the helices. Overall, without the stretching component, pure bending is not an effective mode to open the channel. If excessive bending could occur and couples with significant in-plane stretching [16, 24, 36], then the curvature effect might become important (it is noted that the structural rearrangements found by Meyer et al. [36] are different from our observations [7], since the curvatures of the lipids in these two studies are opposite).

18.4.1.3. Insights of Loading Modes Vs. Mechanotransduction

When an equi-biaxial tension is applied on the membrane, gating is realized primarily by an iris-like expansion of TM1/TM2 helices in the radial direction, as well as tilting of the subunits, whose conformational transitions are directly coupled to the lipid deformation. The S1 pore is also pulled open, in part due to its non-bonded interactions with the transmembrane helices, and in part because of the loop "linkers". The conformational transitions of the intermediate and open structures obtained from the MDeFEM simulation are similar to the structural models [11]. In addition, the simulation results match well with that obtained from all-atom computations [30]. The results demonstrate that the gating process is likely dominated by mechanics principles, including lipid membrane deformation and the deformation/interaction of the helices/loops.

The bending mode is shown to only slightly change in the overall channel radius. Thus, channel gating is relevant to some basic deformation modes (e.g. equi-biaxial tension) but not others (e.g. pure bending). Other deformation modes have also been studied [7] and from the mechanistic point of view, equi-biaxial tension is the most efficient way of achieve full gating. Nevertheless, when these basic modes are combined, such as when bending is coupled with tension, the contribution of bending (membrane curvature) can be important for conformational transitions of the protein.

The present study deals with "external" load acting on the membrane only. In fact, during gating when the hydrophobic residues of the protein are exposed to water, the solvation force that generated "internally" may also play an important role in stabilizing the channel and affecting gating. This important contribution is missing in the present MDeFEM model (see discussions below on future improvements). Thus, the present agreement between MDeFEM and previous experimental and numerical studies may be in part due to error cancellation.

18.4.2. Effects of Structural Motifs

One of the focal points of biophysical studies of mechanotransduction is to understand the diverse roles played by different structural components, such as the helices and loops in MS channels [11, 28]. To that end, one may remove individual group of structural motifs [7] and explore the change of gating behaviors. Since equi-biaxial tension is proved to be the most effective for opening the MS channel, in the following we focus on this basic loading mode. The membrane strain is controlled at 21%, which is required for maximum gating for the full (reference) protein model shown in Figure 18-6a.

In Figure 18-6b [7], after the removal of the loops connecting TM1 and TM2 helices, the constraining effect is reduced and the averaged TM1 tilting angle is decreased by about 10°. The bending curvature of TM1 and TM2 helices also seems to be increased, which further affects the shape of S1 and S3 bundles. Thus, the periplasmic loops are moderately important. Removal of the S3 helix bundle caused insignificant effect on the deformation of the TM1/TM2/S1 helices (Figure 18-6c) [7]; this is because S3 helices are far away from the other protein components. When all structural components are kept except the loops that connect TM1 and S1 helices, the S1 pore becomes distorted, which also affects both TM1 and TM2 helices (Figure 18-6d) [7], illustrating the importance of these loops. These findings are in qualitative agreement with features of the structural model [28].

These trends can also be quantitatively verified in Figures 18-4a and b, where after the S3 helices are removed, both the evolutions of pore radius and the TM1

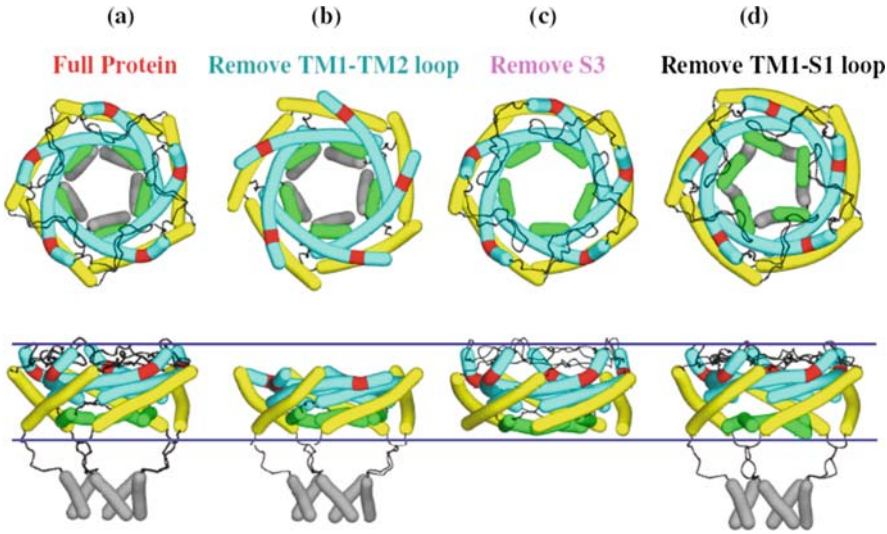


Figure 18-6. Effects of protein structural motifs: (**a**) Full protein (**b**) without TM1-TM2-loop (**c**) without S3 helix bundle (**d**) without TM1-S1 loop. It is found that moderate structural variations are caused by the removal of the loops, whereas the structural conformation is essentially insensitive to the removal of the S3 bundle. Thus, the continuum simulations show that the S3 bundle plays a relatively minor role during the mechanical gating event whereas the loops could constrain gating

helix tilting angle are significantly perturbed (comparing with that of the full protein model); when the loops (either TM1/TM2 or S1/TM1 linker) are taken away, the reduced constraining effect leads to a wider pore in the opened state [7].

The relative inactive role of the cytoplasmic S3 helices show that during the gating, different protein components bear diverse mechanical functions. The loops between TM1-TM2 helices and the loops between TM1-S1 helices, for example, may moderately affect the configuration of the channel in the fully open state. The most important helical components are the transmembrane helices, which directly interact with the membrane and "sense" the forces. These findings [7] are consistent with discussions in previous experimental studies [28], which again illustrate the importance of underlying mechanical principles.

18.4.3. Co-operativity of MS Channels

The biological membrane is highly heterogeneous and rich in proteins and other biomolecules such as polysaccharides. Consider the most fundamental co-operativity problem where two MscLs are in the proximity of each other [7]. The configurations of the channel at membrane strain of 21% are given in Figure 18-7

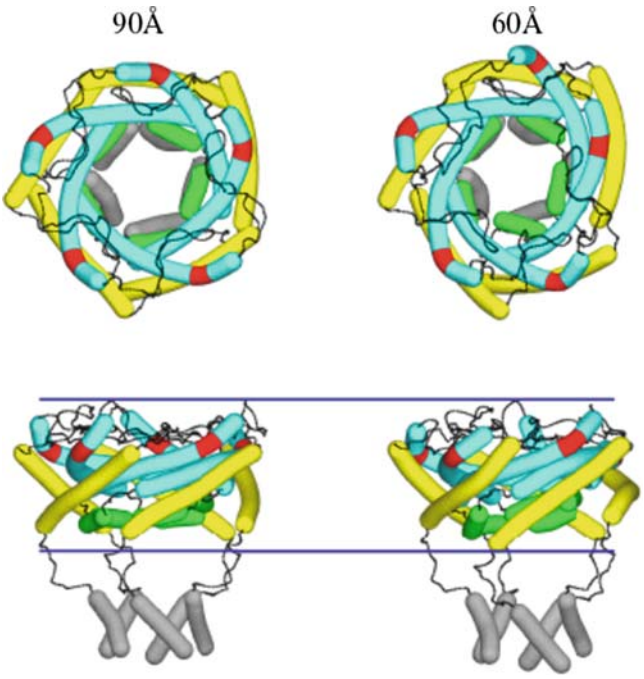


Figure 18-7. The interaction between two *E. coli*-MscLs: the structural configurations of *E. coli*-MscL with the center-to-center separation $\lambda = 90$ and 60 Å (at equi-biaxial membrane strain of 21%). The pore enclosed by TM1 helices becomes increasingly distorted as the channel separation is reduced, indicating stronger magnitude of channel interactions

as the center-to-center separation between the MscLs is varied. When the two proteins are separated far apart, they do not affect each other's conformation. When the separation is reduced to 90 Å, the first column of Figure 18-7 shows that the TM1 pore becomes slightly distorted (with respect to that shown in Figure 18-3) which implies that the MscL starts to sense the existence of its neighbor. When the two channels become closer (with separation 60 Å), the second column of Figure 18-7 demonstrates that the TM1 pore becomes more elliptical and the channel interaction is significant.

In order to determine the critical distance at which co-operativity starts to occur, in Figure 18-8 we plot the bias ratio of the pore (the ratio between the shortest and longest axis) at maximum membrane strain as a function of channel separation. When the two channels are far apart, the bias ratio is 1. The bias ratio decreases gradually with separation, and when the channel center-to-center distance λ falls below about 100 Å, the reduction of the bias ratio is more significant; this leads to a critical separation of about 100 Å [7], which is roughly 4 times the radius of the undeformed lipid cavity (c). This finding is consistent with the thermal dynamics analysis by Ursell et al. [29], which can be explained by lipid elastic deformation: according to the plane stress solution [46], the normalized stress concentration factor of a plate contains two circular hole is increased sharply when the distance between these two

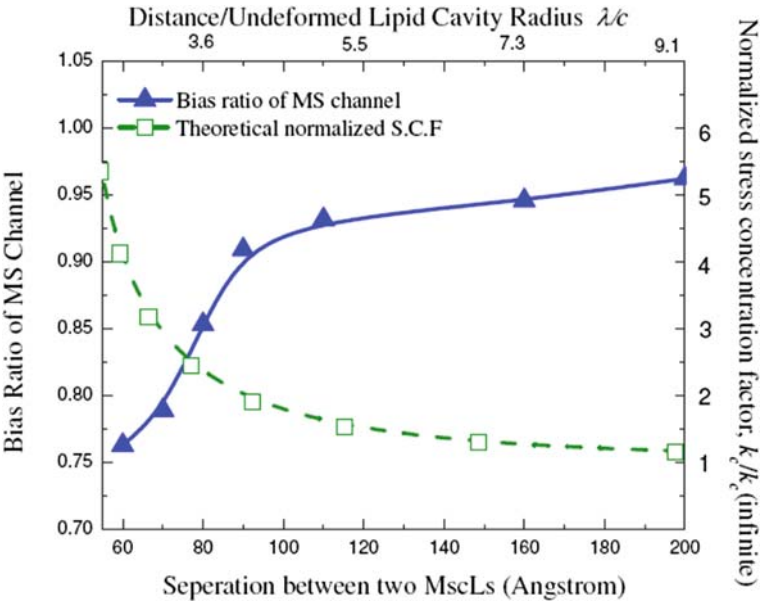


Figure 18-8. Co-operativity between two *E. coli*-MscLs: the bias ratio varies with channel separation, and the results are compared with estimation by elasticity theory (the normalized stress intensity factor, S.C.F., of an elastic sheet containing two circular holes). The bias ratio is the short/long axes ratio of the distorted TM1 pore. The critical separation is about 100 Å below which the two channels strongly interact with each other

holes is below 4 times the hole radius (shown as open-square curve in Figure 18-8). Thus, it is verified again that the mechanical deformation and underlying mechanics principles play a dominant role in MS channel co-operativity.

The above study may provide important insights regarding whether co-operative gating is favored or not. In [29], Ursell et al. used a lipid-centric mechanics model and found that the total membrane energy containing two "open" channels can be reduced when they are close to each other. Note that the protein deformation free energy and structural distortion were not considered in their studies [29], thus whether such co-operative gating is realistic remains to be verified. The MDeFEM-based approach [7] shows that it is possible that the gating threshold is lowered due to co-operativity while the pore shape is distorted; if this is true, a carefully designed channel recording study can be employed to check the current through each channel to explore that with the presence of multiple closely-spaced MS channels, whether the current is smaller due to pore distortion and whether the required energy barrier for gating is reduced.

18.4.4. Large Scale Simulations of Lab Experiments

The advantage of the MDeFEM framework may be demonstrated via the simulation of large scale experiments [7], such as the patch clamp experiment [12, 13, 27] which has historically provided very important information for mechanotransduction. From experiment [12], the geometry of the relatively rigid pipette can be measured (Figure 18-9a) and the opening of the pipette was 1 μm. Frictionless contact is assumed between the vesicle and pipette. Without losing generality, the lipid vesicle may be modeled as an impermeable shell [27] filled with cytoplasm with a bulk modulus that is the same as that of pure water (2.2 GPa) [47]. The averaged size of the liposome used in patch clamp experiment was about 5 μm [27]. To facilitate gating as well as to simplify the analysis, we assume that an *E. coli*-MscL is located at the north pole of the vesicle [7], which would lead to an equi-biaxial stress field and makes the problem axisymmetric.

Figure 18-9a illustrates the undeformed and deformed liposome configurations. By applying a suction pressure, the top portion of the vesicle membrane is attracted into the pipette and forms a bulge shape. The stress is also increased nonlinearly with the suction pressure; at about 0.7 bar of pressure, the local stress near the channel leads to MscL pore opening that is close to full gating (Figure 18-9b). The results also qualitatively agree with the value measured from experiments [7, 12]. The configuration of the final state of the channel is very close to that found in the simulation with equi-biaxial tension on a flat membrane (Figure 18-3), which is expected since the pipette opening is much larger than the protein dimension.

These types of numerical simulations of lab experiments at the cellular scale clearly demonstrate the great value of the multiscale approach. The versatile continuum-based simulation framework may explain, guide, and stimulate new experiments, where the protein location, number, species, and vesicle geometry can be varied. For example, spheroidal or ellipsoidal vesicle geometries may be taken, and the resulting membrane tension stress will depend on the principal curvatures;

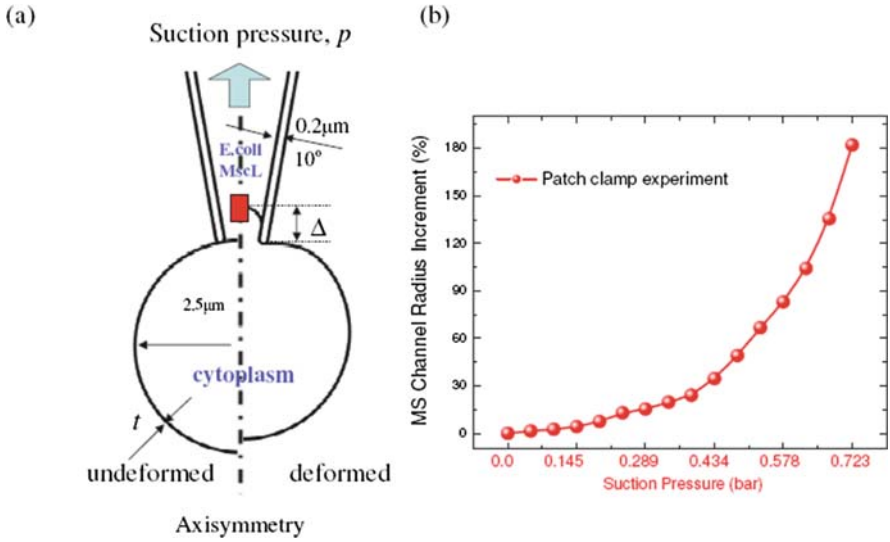


Figure 18-9. Simulation of the patch clamp experiment: (**a**) schematic of the experiment and location of MscL in the lipid vesicle; undeformed and deformed configurations of the lipid vesicle are compared. (**b**) Increment of the effective pore radius as a function of the suction pressure (until full gating)

under such circumstance, the change of MscL locations will lead to different gating behaviors, and the pore is distorted upon non-equi-biaxial stress state. In addition to patch clamp, other experiments such as nanoindentation [7, 48] can be explored as alternative lab tools. In addition, multiple MS channels can be incorporated and the complex stress state under these situations can be very different from the equi-biaxial tension discussed in Section 5.3, since the stress resulted from channel interaction is superimposed with that generated from complex vesicle geometry and loading mode. Refined cytoplasm model can also be undertaken which also affect local stress and strain fields and hence the gating behavior. Eventually, we envision that with continued improvements including those described below, the MDeFEM framework can be readily employed to simulate experiments at a reasonably realistic cellular level that involve MS channels and other membrane proteins.

18.5. FUTURE LOOK AND IMPROVEMENTS OF CONTINUUM FRAMEWORK

For the MDeFEM approach to be quantitatively useful, a number of improvements is needed. In essence, the continuum model can be made increasingly realistic by incorporating more refinements based on atomistic studies. Most improvements outlined below are still based on a MS channel as the model system, although similar insights can be readily applied to a host of other mechanobiology problems.

A more sophisticated model of protein can be adopted along with more heterogeneous mechanical and chemical features. In Figure 18-10a, the molecular

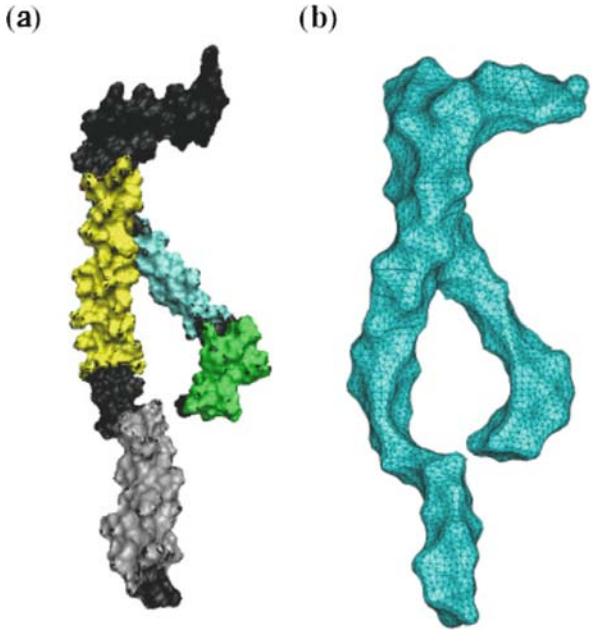


Figure 18-10. A single subunit (TM1/TM2/S1/S2/S3 helices and loops) in *E. coli*-MscL: (**a**) the solvent accessible surface area (SASA) in molecular representation (**b**) the finite element mesh of the continuum model. Such representation is useful for calculating the solvation forces as a further improvement of the continuum-based MDeFEM approach

structure of a subunit chain (containing TM1/TM2/S1/S2/S3 helices and loops) of *E. coli*-MscL is shown. Although the main chain of protein helices can be treated as a cluster of cylindrical cylinders (as in the present MDeFEM model, Figure 18-1), this model makes it difficult to calculate the contribution of solvation forces. With respect to Figure 18-10a, the real helices have irregular atomic "surfaces" dominated by side chain atoms that interact extensively with each other as well as with the environment. A more realistic protein model that reflects its surface topology can be established; an example that uses the solvent accessible surface area (SASA) [49] is shown in Figure 18-10b, similar to the recent study of Bathe [50].

The adaptation of the realistic protein surface morphology enables the calculation of the solvation contribution (electrostatic plus non-polar solvation forces) using popular implicit solvation models (e.g. the Poisson-Boltzmann (PB) [51], Generalized Born (GB) [52], or other implicit methods [53]). The dielectric properties of the protein environment in these calculations can be readily assigned based on the geometry of the elastic membrane (e.g., thickness of the low-dielectric region). It is suspected that the solvation term can cause local lipid membrane bending and stabilize open conformations [44, 45], thereby greatly reduce the required external membrane strain (which was unrealistic in [7]).

Besides incorporating the actual morphologies of the biomolecules, more sophisticated parameterization procedures can also be taken into account. For helices, a critical improvement is to treat them as heterogeneous (e.g. residue sequence-dependent) and anisotropic; although the collective motions of a helix is less sequence dependant, the local deformations can be greatly affected by its heterogeneous and anisotropic properties. Such decoration is essential for describing helix kinking, which is known to play an important role in the gating transition of several channels [54]. For lipid, anisotropic properties are necessary to model the membrane in a more realistic fashion; e.g., the fact that lipid bilayer is unable to bear in-plane shear deformation can be modeled using the transverse isotropic material option. In addition, local residual stress and curvature, as well as hyperelastic and time-dependent viscoelastic features of lipid provide ways to avoid excessive strain and stress observed in simulations using the solid slab membrane model [5]. To parameterize these physical properties, both MD and FEM calculations need to be carried out for different constituents (residues) in helices and (heads and tails) in lipids by either normal model analysis or applying perturbed forces/displacements in different directions. Although these refinements can be quite complicated, they would not significantly compromise the computational efficiency in the mechanical simulations, which are mainly affected by mesh size and density.

The interactions among continuum components can also be treated in a more sophisticated fashion. For example, different interaction parameters can be assigned to hydrophilic and hydrophobic surface regions of protein/lipid. Moreover, the lateral interaction component can be incorporated, which is suspected to be able to stabilize the system and help to create local energy minima (i.e. making certain intermediate states more stable).

Finally, explicit time-dependence in the continuum simulation can be introduced at constant temperature. The temperature may be maintained constant by modifying the equations of motion to include dissipative and zero-mean random force terms, in the spirit of Langevin dynamics, as commonly done in particle based simulations. The frictional coefficients associated with the elements and the distribution of the random force follow the well-known fluctuation-dissipation theorem [55].

18.6. CONCLUSION

In this chapter, we describe a continuum-based approach, the MDeFEM, and illustrate how this framework is particularly useful in the study of gating transitions of *E. coli*-MscL. It is demonstrated that mechanics principles play an important role governing the conformational response of the MS channel to external mechanical perturbations [7]. Based on the results, the advantage, limitation, and future improvements of the continuum-based approach are discussed.

With the experience in applying the MDeFEM approach to MscL and continued efforts of improving the continuum model, similar studies can be made to systems that remain poorly understood, especially those with complex geometry and undergo perturbations inaccessible to conventional all-atom simulations. The multiscale framework can also be applied to contrast the gating mechanisms of

different types of MS channels upon various external stimuli, such as the *Shaker* potassium channel [56, 57] and the MS channel of small conductance (MscS) [45] to identify unique structural and energetic features of mechanosensitive channels. It is envisioned that these studies will lead to further exciting researches to uncover the basic principles of mechanosensations, mechanotransductions and other mechanobiological processes.

ACKNOWLEDGMENT

Supports by National Science Foundation (CMMI-0643726) and National Institutes of Health (R01-GM071428-01) are acknowledged.

REFERENCES

1. Hamill, O.P. and Martinac, B. (2001) Physiological Reviews 81: 685.
2. Ingber, D.E. (2006) FASEB Journal 20: 811.
3. Kamm, R.D. and Kaazempur-Mofrad, M.R. (2004) Mechanics and Chemistry of Biosystems 1: 201.
4. Martinac, B. (2004) Journal of Cell Science 117: 2449.
5. Chen, X., Cui, Q., Yoo, J., Tang, Y., and Yethiraj, A. (2008) Biophysical Journal 95: 563.
6. Tang, Y., Cao, G., Chen, X., Yoo, J., Yethiraj, A., and Cui, Q. (2006) Biophysical Journal 91: 1248.
7. Tang, Y., Yoo, J., Yethiraj, A., Cui, Q., and Chen, X. (2008) Biophysical Journal 95: 581.
8. Gustin, M.C., Zhou, X.L., Martinac, B., and Kung, C. (1988) Science 242: 762.
9. Chang, G., Spencer, R.H., Lee, A.T., Barclay, M.T., and Rees, D.C. (1998) Science 282: 2220.
10. Sukharev, S., Betanzos, M., Chiang, C.S., and Guy, H.R. (2001) Nature 409: 720.
11. Sukharev, S. and Anishkin, A. (2004) Trends in Neurosciences 27: 345.
12. Sukharev, S.I., Sigurdson, W.J., Kung, C., and Sachs, F. (1999) Journal of General Physiology 113: 525.
13. Sukharev, S.I., Blount, P., Martinac, B., Blattner, F.R., and Kung, C. (1994) Nature 368: 265.
14. Wiggins, P. and Philips, R. (2004) Proceedings of the National Academy of Sciences of the United States of America 101: 4071.
15. Kloda, A. and Martinac, B. (2001) Cell Biochemistry and Biophysics 34: 321.
16. Perozo, E., Kloda, A., Cortes, D.M., and Martinac, B. (2002) Nature Structural Biology 9: 696.
17. Elmore, D.E. and Dougherty, D.A. (2003) Biophysical Journal 85: 1512.
18. Gullingsrud, J. and Schulten, K. (2004) Biophysical Journal 86: 3496.
19. Nagle, J.F. and Tristram-Nagle, S. (2000) Biochimica Et Biophysica Acta-Reviews on Biomembranes 1469: 159.
20. Evans, E. and Hochmuth, R. (1978) Topics in Membrane and Transport 10: 1.
21. Turner, M.S. and Sens, P. (2004) Physical Review Letters 93: 118103.
22. Perozo, E. (2006) Nature Reviews Molecular Cell Biology 7: 109.
23. Kung, C. (2005) Nature 436: 647.
24. Wiggins, P. and Phillips, R. (2005) Biophysical Journal 88: 880.
25. Markin, V.S. and Sachs, F. (2004) Biophysical Journal 86: 370A.
26. Martinac, B., Buechner, M., Delcour, A.H., Adler, J., and Kung, C. (1987) Proceedings of the National Academy of Sciences of the United States of America 84: 2297.
27. Sukharev, S.I., Blount, P., Martinac, B., and Kung, C. (1997) Annual Review of Physiology 59: 633.
28. Sukharev, S., Durell, S.R., and Guy, H.R. (2001) Biophysical Journal 81: 917.
29. Ursell, T., Huang, K.C., Peterson, E., and Phillips, R. (2007) Plos Computational Biology 3: 803.
30. Gullingsrud, J. and Schulten, K. (2003) Biophysical Journal 85: 2087.

31. Karplus, M. and McCammon, J.A. (2002) Nature Structural Biology 9: 788.
32. Gullingsrud, J., Kosztin, D., and Schulten, K. (2001) Biophysical Journal 80: 2074.
33. Kong, Y.F., Shen, Y.F., Warth, T.E., and Ma, J.P. (2002) Proceedings of the National Academy of Sciences of the United States of America 99: 5999.
34. Schlitter, J., Engels, M., Kruger, P., Jacoby, E., and Wollmer, A. (1993) Molecular Simulation 10: 291.
35. Bilston, L.E. and Mylvaganam, K. (2002) FEBS Letters 512: 185.
36. Meyer, G.R., Gullingsrud, J., Schulten, K., and Martinac, B. (2006) Biophysical Journal 91: 1630.
37. Lu, G., Tadmor, E.B., and Kaxiras, E. (2006) Physical Review B 73: 024108.
38. Yefimov, S., van der Giessen, E., Onck, P.R., and Marrink, S.J. (2008) Biophysical Journal 94: 2994.
39. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. (1983) Journal of Computational Chemistry 4: 187.
40. Neria, E., Fischer, S., and Karplus, M. (1996) Journal of Chemical Physics 105: 1902.
41. Choe, S. and Sun, S.X. (2005) Journal of Chemical Physics 122: 244912.
42. ABAQUS (2004) Abaqus 6.4 user's manual. 2004: ABAQUS Inc.
43. Gu, L.Q., Liu, W.H., and Martinac, B. (1998) Biophysical Journal 74: 2889.
44. Anishkin, A., Chiang, C.S., and Sukharev, S. (2005) Journal of General Physiology 125: 155.
45. Anishkin, A. and Kung, C. (2005) Current Opinion in Neurobiology 15: 397.
46. Ling, C.B. (1948) Journal of Applied Physics 19: 77.
47. Hartmann, C. and Delgado, A. (2004) PAMM 4: 316.
48. Gordon, V.D., Chen, X., Hutchinson, J.W., Bausch, A.R., Marquez, M., and Weitz, D.A. (2004) Journal of the American Chemical Society 126: 14117.
49. Sanner, M. (2008), http://www.scripps.edu/∼sanner/html/msmshome.html.
50. Bathe, M. (2008) Proteins-Structure Function and Bioinformatics 70: 1595.
51. Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001) Proceedings of the National Academy of Sciences of the United States of America 98: 10037.
52. Feig, M. and Brooks, C.L. (2004) Current Opinion in Structural Biology 14: 217.
53. Zhou, Y.C., Holst, M., and McCammon, J.A. (2008) Journal of Mathematical Analysis and Applications 340: 135.
54. Akitake, B., Anishkin, A., Liu, N., and Sukharev, S. (2007) Nature Structural & Molecular Biology 14: 1141.
55. Zwanzig, R. (2001) Non-Equilibrium Statistical Mechanics. New York: Oxford University Press.
56. Dhaka, A., Viswanath, V., and Patapoutian, A. (2006) Annual Review of Neuroscience 29: 135.
57. Ramsey, I.S., Delling, M., and Clapham, D.E. (2006) Annual Review of Physiology 68: 619.

CHAPTER 19

OUT OF MANY, ONE: MODELING SCHEMES FOR BIOPOLYMER AND BIOFIBRIL NETWORKS

E.A. SANDER[1], A.M. STEIN[2], M.J. SWICKRATH[1], AND V.H. BAROCAS[1]

[1] *Department of Biomedical Engineering, University of Minnesota, Minneapolis, MN 55455, USA, e-mail: sande399@umn.edu, baroc001@umn.edu*

[2] *Institute for Mathematics and Its Applications, Minneapolis, MN 55455, USA*

Correspondance to V.H. Barocas, Department of Biomedical Engineering, University of Minnesota, 7-105 Hasselmo Hall, Minneapolis, MN 55455, USA, baroc001@umn.edu

Abstract: Tissues are structurally and compositionally complex materials that must function in a coordinated fashion at multiple length scales. Many of the structural proteins in soft tissues and in cells form biopolymer networks that provide mechanical benefits and coordinate cell-directed physiological activities. Complicated phenomena operate at multiple scales and are governed to varying degrees by the properties of networks; thus, mechanical models are a necessary tool to unravel the relationships among individual network components and to determine the aggregate properties and functions of cells and tissues. In this work, we review major biopolymers, their function, and the general mechanical behavior of biopolymer gels. We then discuss some network imaging techniques and methods for constructing and modeling networks /in silico/ – including multi-scale methods. Finally, we return to the specific biopolymers, including actin, microtubules, intermediate filaments, spectrin, collagen I and IV, laminin, fibronectin, and fibrin, and discuss what has been learned from the different models. Biopolymer network models, especially when combined with ever-improving experimental methods, have the potential to answer many fundamental questions in mechanobiology

Keywords: Collagen, Actin, Multi-scale

19.1. INTRODUCTION

Tissues are complex materials composed of cells and extracellular matrix (ECM) proteins that must function in a coordinated fashion at multiple length scales. Biopolymer networks that form the ECM vary in composition and organization in a manner that confers suitable mechanical properties to the tissue and allows tissues to function in their physiological capacity. Mechanical loads and constraints applied to the whole tissue are transmitted down through the matrix and into the cells. The cells, which are stabilized and detect mechanical forces through the cytoskeleton – an intracellular network – then respond through a variety of dynamic activities that can lead to growth, remodeling, and adaptation.

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 557–602.
DOI 10.1007/978-1-4020-9785-0_19,

The network architecture provides many beneficial properties to cells and tissues. Networks produce strong and stable structures with a minimum investment in materials. Their open configuration enables various transport processes (e.g. nutrient diffusion) to occur with less hindrance, and permits cell locomotion when appropriate (e.g. white blood cells). Also intrinsic to networks is the ability to communicate signals rapidly and at a distance. In terms of mechanical signaling, such a system may provide a means to coordinate cell behavior within tissues [1, 2].

With many complicated phenomena operating at multiple scales and governed to varying degrees by the properties of networks, mechanical models become a necessary tool for unraveling the relationships between individual network components and the aggregate properties and function of cells and tissues. *A sampling of the kinds of questions about cell and tissue function that mechanical models can provide answers to includes:*

- How do tissue material parameters depend on the biopolymers and networks structure, e.g. what components and deformations contribute to the elastic response and where does time-dependent viscous behavior come from?
- How does the ECM microstructure reorganize to accommodate macroscopic strain? Do fibers rotate, stretch, bend, or buckle?
- What do those rearrangements mean in terms of mechanical signals that can be sensed by cells?
- How does a cell sense mechanical force and translate that into a decision to act in a certain way? For example, how do stimuli cause synthesis or degradation?
- How do mechanical changes in the ECM lead to different diseases, and how might they be prevented?
- What components and what structural arrangements are necessary to produce a functional engineered-tissue?

The questions listed above are all inherently *multiscale*, with the tissue scale ($\sim 10^{-3}$ m), cell scale ($\sim 10^{-5}$ m), ECM fiber scale ($\sim 10^{-7}$ m), biomacromolecular scale ($\sim 10^{-9}$–10^{-8} m), and atomic scale ($\sim 10^{-10}$ m). Although not all scales necessarily need to be studied to answer every question, any approach to understanding mechanobiology and biomechanics from a structural standpoint must respect their scale-spanning nature.

In this work, which emphasizes network mechanics, we focus on the transitions from biopolymer to cell (as in the cytoskeleton), biopolymer to tissue (as in basement membrane), and fiber to tissue (as in ECM or bioartificial tissues). In each case, the network consists of long, thin units connected in a structured or unstructured manner.

In the next section, we briefly review the major biopolymers and their function, followed by some general mechanical properties of biopolymer gels. We then discuss various methods to analyze images of networks and construct computer models there of. Next, we discuss the current approaches to network modeling in the general case and different methods that have emerged. Finally, we return to the specific biopolymers and discuss what has been learned from the different models.

19.2. BIOPOLYMERS OF INTEREST

Cells interact with the physical world, and that interaction depends on networks inside and outside the cell. Inside the cell, a network of actin filaments, microtubules, intermediate filaments, and other proteins come together to form the cytoskeleton. This ensemble of intracellular proteins stabilizes the cell structure and plays a role in many cellular phenomena, including changes in cell shape and cell division. In addition, the cytoskeleton appears to play a prominent role in translating environmental cues, both mechanical and chemical, into a cellular response, which may take the form of biosynthetic activity [3–6] or even programmed cell death (apoptosis) [7, 8].

Outside the cell, a network of proteins, most commonly with collagen as the backbone, forms the extracellular matrix. The ECM composition and organization confers functionality to a tissue and provides a conduit for mechanical signals to alter cellular response. In many tissues, the ECM is in a constant state of turnover. The cells in the host tissue respond to changes in the microenvironment by degrading and synthesizing ECM proteins. Such changes can lead to growth and adaptation, e.g. tissue growth with exercise [9], or can lead to disease, for example glaucomatous damage to the optic nerve head [10] or hypertensive arterial wall thickening [11]. Other-ECM related diseases are congenital and result in impaired tissue function with devastating consequences. In Alport's syndrome, for example, the genes encoding for a crucial component of the basement membrane malfunction [12]. The basement membrane structure is altered, which greatly impairs the molecular sieve structure of the kidney glomerulus, making it vulnerable to high pressures and more susceptible to proteolytic attack. Consequently, understanding the interplay between molecular interactions and macroscopic tissue mechanics is crucial to understanding many pathologies.

In this section, we introduce and briefly describe some of the monomers that form key intracellular and extracellular networks. The interested reader should consult the references listed in Table 19-1 for a more comprehensive review on each protein.

19.2.1. Intracellular Networks

19.2.1.1. Actin

Actin filaments (F-actin) are composed of a linear chain of G-actin subunits that are constantly and dynamically added to or removed from the ends of F-actin in a manner dependent on the local G-actin concentration. This process enables the cell to reorganize the cytoskeleton, migrate, attach to a substrate, and respond to signaling [13–17]. The G-actin subunits, which are approximately 2–3 nm in diameter [18] form semiflexible F-actin filaments that are approximately 5–7 nm in diameter and can ultimately assemble into hierarchical bundles and networks that span the interior of the cell. For a single actin filament, the stretching stiffness is $K_s = 4.4\times10^{-8}$ N [19], the bending stiffness is $K_b = 7.3\times10^{-26}$ Nm^2, and the persistence length (defined later) is $l_p = 17$ mm [20]. Actin molecules associate readily with divalent cations (Mg^{2+} and Ca^{2+} in particular) giving the molecule the capacity to

Table 19-1 Intracellular and extracellular biopolymers

Type	Approximate dimensions	Macromolecular structure	Mechanical function	Reference
Intracellular				
Actin	5 nm dia, 10–20 μm length	Bundles and networks	Cytoskeletal tension component	[14]
Microtubules	25 nm dia, 10–20 μm length	Hollow tubes emanating from MTOC[a]	Cytoskeletal compression component	[231]
Intermediate filaments	Variable, between 5 and 25 nm dia	Filaments coupled to cytoskeletal junctions	Cytoskeletal junctional component	[28]
Spectrin	200 nm long head-to-tail tetramers	Intertriangulated network with Actin	Cytoskeletal compression component	[232]
Extracellular				
Collagen I	~300 nm long, 1.5 nm dia[b], assembles into higher order fiber structures	Fibers	Extracellular matrix component	[33]
Collagen IV	800 nm long hexameric units	Polygonal network	Extracellular matrix component	[12]
Laminin	Cruciform, 75 nm wide, 115 nm long	Ionically cross-linked network	Extracellular matrix component	[50]
Fibronectin	2–3 nm dia, 60–70 nm length	Fibrillar network	Extra- to intracellular mechano-transduction	[52]
Fibrin	Fibrinogen monomer 5–7 nm dia, 45 nm length, assembles into larger fibrin fibers upto hundreds of nm in dia	Branched network	Clot formation	[56, 233]

[a]MTOC (microtubule organizing center).
[b]Tropocollagen.

complex with ADP and ATP. The conversion of ATP to ADP via hydrolysis through the ATPase myosin results in a conformational change in the F-actin molecule. This mechanochemical phenomenon driven by myosin has led to the colloquial reference of myosin as a motor protein and gives F-actin the capability to induce mechanical forces within the interior of a cell [21–23]. Furthermore, the protein ARP23 causes branching of actin filaments, contributing to the network structure [24].

19.2.1.2. Microtubules

Microtubules are another major cytoskeletal component critical to cell function. They are created when tubulin, a heterodimer of α-tubulin and β-tubulin, polymerizes to form stiff hollow tubes ~25 nm in diameter. Microtubules are also controlled

by the polymerization/depolymerization of their subunits. Microtubules are involved in a number of cellular processes including vesicle transport and cell division [25]. Their rigidity helps support organelles and maintain cell shape. Microtubules may also oppose the tensile forces generated by F-actin [26, 27].

19.2.1.3. Intermediate Filaments

Intermediate filaments (IFs) comprise a third classification of cytoskeletal components that are more stable structures than F-actin and microtubules [28]. IFs, which are ~10 nm in diameter, are "intermediate" in size when compared to F-actin and microtubules. Intermediate filaments can be found at the transcellular junctions (e.g. gap junctions, tight junctions, desmosomes and adherens junctions) as well as at anchoring plaques to the extracellular matrix (e.g. focal adhesions and hemidesmosomes). IFs are also linked to F-actin on the interior of a cell creating a pathway for the mechanotransduction of extracellular mechanical phenomena.

19.2.1.4. Spectrin

Spectrin, a cytoskeletal component specific to the red blood cell, is composed of a dimer of either α-spectrin or β-spectrin, both of which are ~250 kDa. The dimers arrange in an anti-parallel arrangement forming tetramers that associate with short actin filaments (~15 subunits) creating an inter-triangulated actin-spectrin network conferring mechanical stability and enabling a blood cell to compress and subsequently expand [29, 30]. The spectrin-actin network is intrinsically important to the transport of erythrocytes, allowing the erythrocyte to modulate shape as it passes through narrow capillaries [31].

19.2.2. Extracellular Networks

The ECM functions as a support and anchoring structure for cells and as a means of tissue compartmentalization. The following components represent the major network-forming ECM molecules, and include type I collagen, type IV collagen, laminin, fibrin and fibronectin.

19.2.2.1. Collagen I

Collagen – the most abundant protein in the body – refers to a family of structurally and functionally related proteins that consist of three helically wrapped polypeptide chains [32, 33]. Type I collagen is a fibrillar collagen and accounts for 90% of all collagen (other fibrillar collagens include types II, III, V, IX). The fundamental unit of collagen, tropocollagen, is 280 nm in length and 1.5 nm in diameter [34]. Tropocollagen is composed of three polypeptide chains, or α-chains, that wrap around each other to form a right-handed triple helix. Tropocollagen is secreted into the ECM where it is modified enzymatically, assembled into quarter-staggered subfibrils, and covalently cross-linked [35]. Subfibrils associate laterally into fibrils and fibers, a distinction based mainly on size. Fibril diameters range from 10 nm to several hundred nm. They can organize into higher-order fibril bundles or fibers that

can be hundreds of nanometers in diameter [36] and hundreds of micrometers in length [37]. For collagen-I, the properties of the triple helical monomer have been measured to be $K_s = 5.08\times10^{-10}$ N, $K_b = 3.36\times10^{-37}$ N-m^2, and a persistence length $l_p = 14.5$ nm [38]. In aqueous conditions, collagen fibers have been reported to have Young's Moduli ranging from 32 to 900 MPa [39–42]. The bending stiffness for native collagen fibers ranges from 3×10^{-15} N-m^2 to 6×10^{-15} N-m^2 [43]. Many collagen fibers are heterotypic, meaning they are composed of more than one type of collagen [44]. Due to the complexity of native tissues, reconstituted collagen gels have served as simple but important in vitro tissue models [45–47].

19.2.2.2. Collagen IV

Unlike the fibrillar collagens, type IV collagen assembles into a mesh-like network that serves as the scaffolding for basement membranes. Basement membranes anchor and support endothelial and epithelial cells to connective tissue and provide physical barriers that allow for tissue compartmentalization. Collagen IV is comprised of three polypeptide chains associated as a triple helix and measuring approximately 400 nm in length [12, 48]. Six genetically distinct type IV collagen chains exist and are denoted as α_1–α_6. The chains assemble specifically forming three heterotrimers [α_1(IV)]$_2$,α_2(IV), α_3(IV), α_4(IV),α_5(IV), and α_5(IV)]$_2$,α_6(IV)]. The relative concentration of each heterotrimer is dependent upon the tissue and the functional requirements of the collagen IV network. Type IV collagen interacts cooperatively with a variety of proteins and glycoproteins in forming the membrane [49]. Additionally, collagen IV can be reconstituted in vitro [49].

19.2.2.3. Laminin

Laminin has many functional roles in the ECM that relate primarily to cell attachment, including induction and maintenance of cell polarity, establishment of tissue barriers and compartments, organization of cells into tissues, and prohibition of attachment-induced cell death [50]. Laminin is a cross-shaped heterotrimer of glycoproteins comprised of several combinations of α, β, and γ subunits resulting in 15 distinct heterotrimers. In general, the molecule is comprised of three short arms of ~37 nm and a long arm of ~77 nm. All ends of the molecule have a globular domain providing functionality. In vitro, laminin can aggregate into networks in a concentration-dependent and thermally-reversible manner in the presence of divalent ions such as Ca^{2+} and Mg^{2+} [48].

19.2.2.4. Fibronectin

Fibronectin (FN) is a cell-secreted, soluble dimer, which polymerizes into an insoluble fibrillar network that facilitates cell attachment to the ECM (collagen types I–III and V, in particular) [51]. The fibronectin subunit, a dimer of polypeptide subunits 60–70 nm in length and 2–3 nm in diameter, associate into dimers that interact directly with integrins on a cell surface [22, 52]. Cytoskeletal tension created across the integrin stretches the FN dimer and exposes FN-FN binding sites to other

interstitial FN dimers, resulting in FN fibril and network formation. Because FN directly attaches to the cytoskeleton, and the FN network is assembled by mechanical mediation, it is thought that FN plays an important role in influencing cell shape, organization, and locomotion. Additionally, FN deposition is observed in a wide variety of wound healing processes, usually preceding the deposition of a more permanent collagen matrix [53].

19.2.2.5. Fibrin

Fibrin networks form during blood clotting as part of the wound healing process. Fibrin is formed from the assembly of fibrinogen, a trinodular ~340 kDa protein present in plasma that is 45 nm in total length [54]. Polymerization is catalyzed by thrombin, which enzymatically cleaves the N-termini of the α and β chains creating the "A" and "B" polymerization sites, respectively. The fibrin monomers arrange in a half-staggered arrangement aligning complimentary bonding sites creating oligomers that arrange in pairs creating dual-stranded protofibrils. Protofibrils associate laterally-forming fibers which ultimately aggregate, constrained by the ionic conditions, into bundles with a paracrystalline structure and distinctive banded pattern [55]. The bundles undergo branching creating a three-dimensional network that is covalently cross-linked by factor VIIIa concomitantly with the release of fibrinopeptide B [56]. In addition to its in vivo function, fibrin has also emerged as an attractive scaffold for tissue engineering [57, 58].

19.2.3. The Mechanical Behavior of Biopolymers

The biopolymers above can be examined in their native state, but frequently are purified and reconstituted in gel form. The in vitro gel is a much simpler system than cells and tissues, while still providing many of the rich mechanical properties observed in the in vivo systems. The bulk properties of gels are frequently measured with a rheometer. A common test performed involves casting a biopolymer gel between two surfaces, often parallel plates, and oscillating one plate back and forth at frequency ω while the other is held fixed. In the small strain limit of a linear viscoelastic material, the stress-strain response is given by

$$\sigma(\omega,t) = G'(\omega)\gamma(\omega t) + G''(\omega)\frac{\dot{\gamma}(\omega t)}{\omega}, \tag{19-1}$$

where σ denotes the stress, γ denotes the strain, G' denotes the elastic modulus and G'' denotes the loss modulus. For a perfectly elastic material $G'' = 0$, and for a perfectly viscous material $G' = 0$. For typical biopolymers such as actin and collagen, the elastic character of the gel dominates at low frequencies (less than 100 Hz) – G' is an order of magnitude greater than G'' [59, 60] – thus making it possible to measure the elastic properties of the network with these tests. The viscoelastic character of the gels stems from molecular-level rearrangements and fluid-solid interactions (poroelasticity). The viscoelastic behavior is important but will not be discussed here. For more information see [47, 61–65].

The gel's elastic character depends on the biopolymer concentration c and the cross-links formed. One typically observes power-law scaling of the form $G' \sim c^x$, where for actin $x = 1.4$ when no cross-linker is present [66], and $x = 2.5$ in the presence of very strong and stiff cross-linkers [60]. Another parameter of significance is the ratio of cross-linker to polymer concentration, R. Again, there is typically power-law scaling of the form $G' \sim R^y$, but a transition point exists, where at small $R, y = 0.1$, whereas at larger R,y may range from 0.4 [67] to 2.0 [68], depending on the stiffness of the cross-linker. Recently, it has also been observed that when sheared, these materials tend to compress and pull the shear plates together [69].

At small strains, typically less than 10%, it is reasonable to treat biopolymer gels as linear viscoelastic materials. At larger strains, however, they typically stiffen, with a modulus that can increase by 2–3 orders of magnitude [70]. A number of models give alternative explanation for this stiffening. At very large strains, the material breaks, undergoing irreversible deformation. Some materials, including actin, fail earlier at increasing densities [60] while other materials, like collagen, break at the same strain regardless of the density [71]. These mechanical properties are summarized in Figures 19-1, 19-2 and 19-3.

In addition to shear, extension and compression tests have also been conducted on biopolymer networks (mostly collagen) – for a review see [47]. In uniaxial extension, unconstrained network fibers rotate and align with the displacement axis before the network stiffens due to fiber resistance to axial

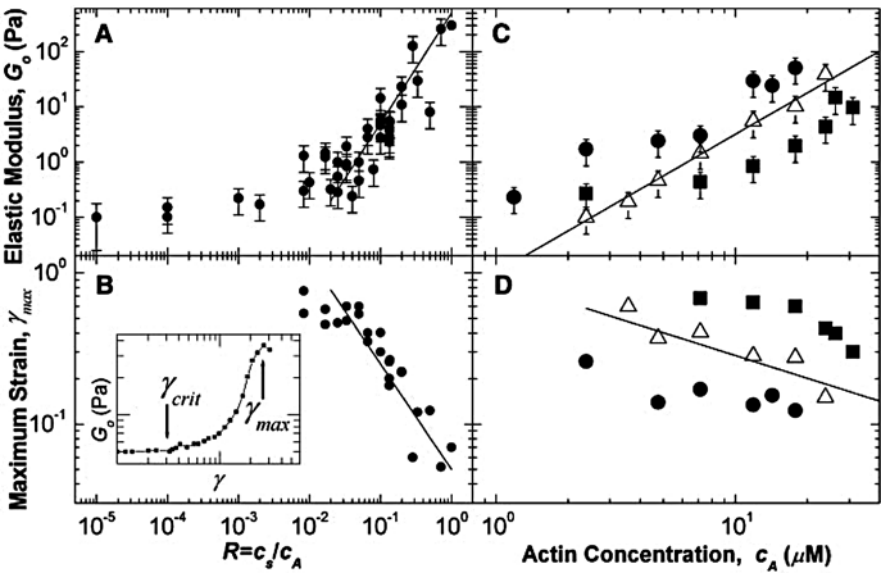


Figure 19-1. G' Properties of actin from. The elastic modulus (G') and maximum strain ($\gamma_{\max}$) for actin networks cross-linked with scurin, as a function of actin concentration (c_{A}) and cross-linker/actin ratio (R). From [204] with permission

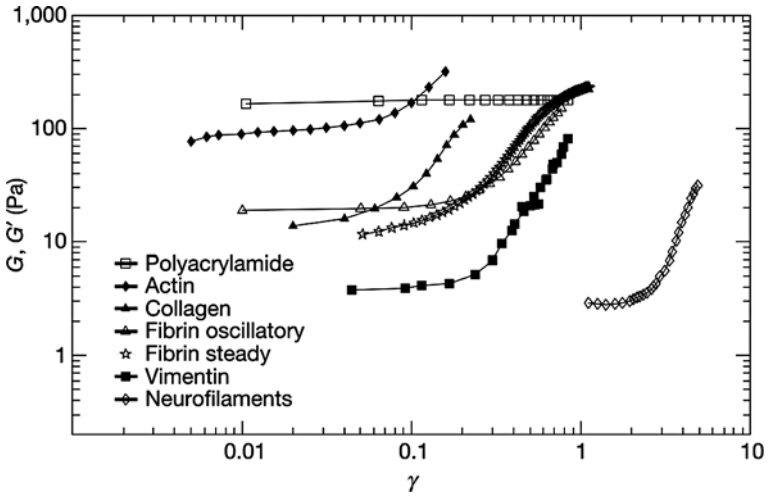


Figure 19-2. Strain stiffening properties of biopolymers. From [70] with permission

stretch [72–75] thereby producing the characteristic J-shaped stress-strain curve observed in soft tissues [71, 76, 77]. The rapidity of stiffening is dependent on network properties and constraints. For example, a more cross-linked network stiffens faster at a lower extension, as does a network that is constrained

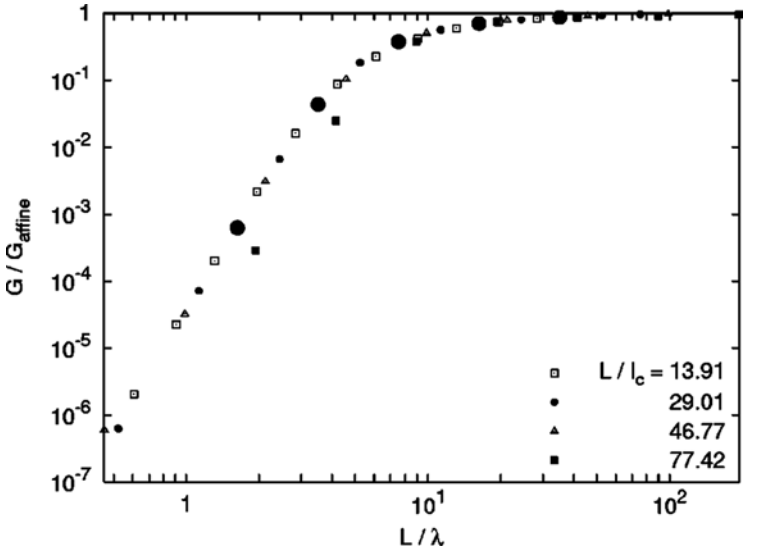


Figure 19-3. Nonaffine-affine transition in mechanical beam networks. G′ scaling with L/λ, where $\lambda = l_c(l_c/l_b)^{1/3}$. Note that when the spacing between cross-links is large, the network is much less stiff than the nonaffine network. (Reprinted with permission from Head et al. [143]. Copyright (2003) by the American Physical Society. http://prola.aps.org/abstract/PRE/v68/i6/e061907)

transversely from contracting inward when stretched. Soft tissues and collagen gels exhibit a reduction in the peak stress and the amount of hysteresis between the loading/unloading curves that converges to a stable value when the stretch protocol is repeated at the same rate and to the same extent (preconditioning). It has been suggested that this behavior is due to microstructural rearrangements, although the specific cause remains unknown. One possibility is that non-covalent interactions between fibers continue to break as a result of the cyclic stretch until a stable configuration is reached.

The gel response to compression testing is more complicated, and the dissipation mechanisms involve molecular interactions and interstitial flow. Gels are often confined laterally in a chamber and compressed with a porous piston to allow fluid flow out of the gel [65, 78], an experiment inspired by the articular cartilage community [79, 80]. In collagen gels, the network response was found dependent on the time scale of the deformation, with step and ramp tests resulting in fiber collapse near the piston or fiber bending that induced network restructuring throughout the gel [62].

Higher collagen concentration generally translates to better mechanical properties, but again it is unclear whether the underlying cause is more cross-links, larger fibers, or other changes to the network architecture. It is well known that collagen fiber and network architecture is highly dependent on the gelation conditions, including pH, temperature, and ionic concentration [71, 81, 82]. In the absence of cross-linker, it has been found that the storage modulus scales with collagen density c_c by $G' \sim c_c^{(2.45\pm0.25)}$ [83, 84]. This is markedly different from the $G' \sim c_a^{1.4}$ for uncrosslinked actin, which suggests that even in the absence of chemical cross-linkers, the fibers naturally cross-link, a conclusion in agreement with macroscale observations as well [62].

As already noted, tissues are compositionally and architecturally more complex than single-phase biopolymer networks. As a result, other ECM components, including proteoglycans, elastin, laminin, and fibronectin, have been added to collagen gels in order to assess their impact on tissue mechanics [85, 86]. In general, the changes in G' and G'' were concentration dependent with the additives either aggregating collagen fibers together, as was the case with the proteoglycans studied, or thickening the fibers by coating them. Regardless of the macromolecule added, interpreting the results of such experiments is difficult because it is not known how the proteins affect network assembly or how the resulting structure compares to native tissues.

Nevertheless, these studies are relevant to understanding how tissues are built, particularly skin and cartilage, which share similar structural arrangements. Tissues like tendons, on the other hand, are highly organized and cross-linked into a hierarchical structure designed to resist high tensile load and store energy and do not behave like collagen gels. A discussion on the mechanical properties of soft tissues is quite involved and beyond the scope of this review. The interested reader is referred to the works of Fung [76] and Humphrey [87] for more information.

19.3. NETWORK IMAGING, EXTRACTION, AND GENERATION

19.3.1. Imaging

19.3.1.1. Fiber-Level Imaging

To produce an accurate representation of a biopolymer network, one must first obtain images of the microstructure. The difficulties involved are many, and each imaging technique has its advantages and disadvantages. The most common means for obtaining microstructural information relies on light level histology techniques [34]. Different colored dyes or stains are applied to thin, fixed sections of tissue to visualize the different matrix components. Histology is relatively inexpensive and easy to do, and it can provide spatial information for multiple species. Histology's main detractions are that it is labor intensive and only moderate in resolution (submicron). In addition, the information obtained is two-dimensional and prone to artifact, which can arise during fixation, dehydration, sectioning, or staining. More specific staining can be achieved, often through the use of fluorescent antibodies that bind specifically to the target molecule. When the sample is illuminated with a specific bandwidth of light, only the tagged molecules are imaged. Serial sections through a sample can be reconstructed into 3d datasets of the tissue's microstructure [88, 89], or to obtain fiber orientation [90]. These methods are again labor intensive and often produce artifacts, and the 3D reconstructions are computationally demanding. More importantly, the real-time microstructural response to macroscopic loads is not accessible.

More advanced imaging technologies have emerged, which are also capable of providing 3D data sets. Magnetic resonance imaging (MRI) [91], computed tomography (CT) and micro CT (μCT) [92–94], and optical coherence tomography (OCT) [95–98] have all been applied to tissues, most notably bone. These techniques allow 3-D imaging of living tissues, but are limited in their ability to identify differences in soft tissues and do not provide sufficient resolution to image at the scale of the microstructure [99].

In the case of purified gels, confocal microscopy [75, 100–102], and multiphoton microscopy [82] can be used obtain 3-D images of the networks without destroying their network architecture. Typically, the point spread function of the system is on the order of 500 nm, whereas for collagen, the fiber radius is often less than 100 nm [83]. Thus while fibers of small diameters are visible, the precise radii of the fibers and details of the fibril architecture cannot be resolved.

Electron microscopy (EM) provides a range of techniques useful for visualizing the microstructure in detail because resolution is on the nm scale. EM has been used to directly observe a variety of biomolecules, including type IV collagen [103], laminin [104], and spectrin [105]. Scanning electron microscopy (SEM) is often used because of its superior depth of field, which also has the disadvantage that quantification of fiber dimensions is difficult without resorting to stereoscopic techniques. Transmission electron microscopy (TEM) is much more suited to quantitative measurements because the sample is sectioned into thin slices. TEM can also be combined with other preparatory techniques, such as quick-freeze/deep

etch, where a replica coating of the sample is imaged instead of the sample itself [106, 107]. Sample preparation in EM is also difficult and sample artifacts similar to those that occur in light level techniques are also present. Because of the higher resolution, however, artifacts are magnified and present greater difficulties in extracting the true microstructure. Cryo-(SEM) presents a more pristine picture than conventional SEM because water-associated structures are not dehydrated. In this technique, the sample is rapidly frozen so that the water vitrifies [108, 109]. 3-D reconstructions are also possible using electron tomography, which has been used to reconstruct cell structures [110] and single molecules, including collagen and fibrillin [111].

19.3.1.2. Indirect (Population-Level) Imaging

Non-invasive, indirect measurements of the fiber microstructure are possible by probing the sample's optical properties. Techniques, such as small-angle light scattering (SALS) [112] and polarimetric fiber alignment imaging (PFAI) [72, 113] do not image the fibers directly, but instead make quantitative measurements of the fiber population based on the optical properties of submicroscopic fiber networks. In SALS, the pattern of scattered laser light transmitted through a sample provides a local fiber orientation distribution, which can be used to generate dynamic alignment maps during mechanical testing of tissues [114]. PFAI exploits the birefringent properties of the fiber and the difference in refractive index between fiber and solution to assess principal fiber direction and degree of alignment by measuring the change in amplitude and phase of the elliptically polarized light transmitted through the sample. Consequently, it can only be employed if the biopolymer network is birefringent, as is the case with collagen and fibrin. PFAI has been used extensively to generate 2-D network alignment maps in a variety of static [72, 115, 116] and dynamic [62, 117, 118] bioartificial tissue systems. Indirect techniques do not have the resolution that confocal methods have but they are easier to implement and can survey whole tissue samples under a variety of loading schemes. One limitation, however, is that the samples must be sufficiently transparent (i.e. thin enough). Otherwise sample sectioning or optically clearing the tissue with a hyperosmotic solution may be required [114]. Another issue is that neither method can discriminate between different fiber populations. For example, a fiber orientation distribution obtained from a remodeled fibrin gel cannot distinguish fibrin fibers from newly formed collagen fibers. Regardless of the real-time imaging method used, gaps between scales still exist which can only be addressed with multi-scale computational models and a cohort of imaging techniques.

19.3.2. Network Extraction

Another issue is how to describe the microstructure once an image or representation of the microstructure is obtained. Morphometric and stereologic methods have often been employed to describe tissue microstructure [119, 120]. These descriptors can provide exact quantities, such as volume fraction and number of objects, or

distributions, such as fiber length, width, and orientation angle. Several tensor representations have been employed to describe material anisotropy and micro-structural alignment [121, 123]. They can be constructed from image-based measurements, such as mean intercept length [119, 124] or from Fourier transform methods (FTM) [125], and are convenient for capturing the principal direction and strength of fiber alignment.

Other image processing techniques have also been used to extract fiber and network features, as well as to map myocardial fiber orientation [126] and to quantify cytoskeletal reorganization in response to shear [127], stretch [128], and wound healing [129]. Some of these methods involve first thresholding the intensity image into a binary image. Additional processing might include the use of filters for edge detection and gradient calculation [126], or skeletonization and tracking [130] to determine fiber orientation and magnitude, or to reconstruct the network [83, 131, 132]. In addition, Fourier methods [125, 129, 133, 134] and the Hough transform [135, 136] have proven useful for obtaining fiber distributions. The majority of these methods have been developed for 2-D images, and some have been extended to 3-D [83, 131, 132, 137] For more information on image processing techniques see Gonzalez et al. [138].

19.3.3. Model Network Generation

A variety of methods have been implemented to create networks, not all based on measurements of the microstructure. The simplest network model assumes an idealized geometry representative of the material, such as a hexagonal cellular solid unit cell [139]. Another possibility is to use an established algorithm, such as Voronoi tessellation or Delauney triangulation, to subdivide a region into a mesh. Methods of this type are useful but generate more ordered, cellular-solid-like networks, which only share some features with fibrous networks [140, 141]. Consequently, one should consider whether the material to be modeled is more appropriately described as a cellular solid or fiber network when creating the network geometry (Figure 19-4).

The generation of random 2-D straight-fiber networks, known as a Mikado model, involves randomly selecting network properties from a distribution function (e.g. uniform, von Mises, etc.). Typically, fiber position and angle are selected randomly, and locations where fibers intersect are made into cross-links [141–144]. Other distributed networks properties can be used to shape the network including fiber length and aspect ratio [145, 146]. In some cases the networks created are periodic, meaning that fibers that overlap the box boundaries are made to wrap back around to the other side [145, 147]. Periodic boundaries are generally used to remove end effects and to more easily impose network properties, such as total fiber length and network volume fraction [145].

These techniques can be extended to generate 3-D networks but an additional angle (ranging from 0 to π) must be included. Automatic 3-D skeletonization algorithms have been developed for extracting the network structure [83, 131, 132] from

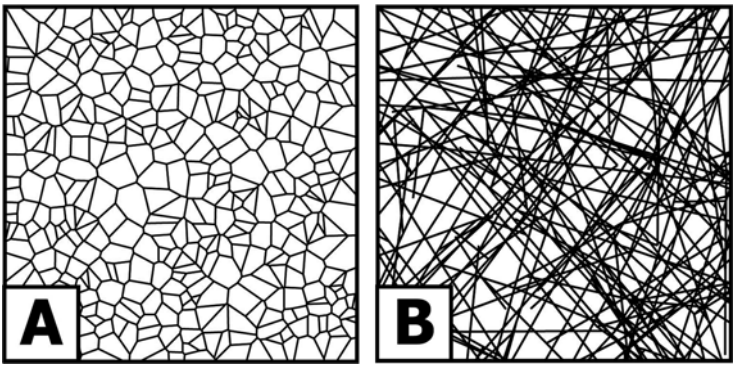


Figure 19-4. Cellular vs. Fibrous Networks. (**a**) A random cellular network produced via Voronoi tessellation using random points. (**b**) A random fiber network created from a random growth algorithm. See Huessinger [139] for details on the differences

images and recently, Stein et al. [83] have validated that the architectures extracted by these algorithms have realistic geometric and mechanical properties.

Most of our work has focused on the mechanics of biopolymer gels and tissue equivalents [74, 117, 148–150]. In these studies, 3-D fiber networks were created in a stochastic process that resembles the process of collagen fiber formation in gels. First, a number of seed points are generated inside of a box. A fiber grows bidirectionally from each seed point until intersecting the boundary or another fiber. We have recently set up the method to generate statistically equivalent networks to those obtained from polarized light imaging by adjusting the random direction distribution and checking to match the observed structure [117].

19.3.4. Network Generation via Energy Minimization

Networks can also be generated through energy minimization techniques such as the Metropolis-Hastings (MH) importance sampling algorithm [151, 152] (a Monte Carlo technique). MH is used to generate the structure and interactions of dynamic chemical systems from time-independent and stochastic rules [153–155]. If the rules of the simulation are posed adequately, two differing and commensurate sequences of random numbers should generate statistically equivalent results (i.e. the results will agree to within a small "statistical error"). Consequently, the MH algorithm is a powerful tool to bridge how nanoscale chemical energetics yield macroscopic networks with determinable mechanical properties.

The underlying principle of the MH algorithm is to calculate a thermodynamic minimal average energy $\langle U \rangle$, of an ensemble of m molecules, $\{n_1,...,n_m\}$, at a given temperature, T, using the following equations:

$$\langle U \rangle = \frac{1}{Q} \int U(n_1,...,n_m) \exp\left[-U(n_1,...,n_m)/kT\right] dn_1...dn_m, \tag{19-2}$$

$$Q = \int \exp\left[-U(n_1,...,n_m)/kT\right] dn_1...dn_m, \tag{19-3}$$

where k is the Boltzmann constant. The difficulty in performing such a calculation is that the normalizing quotient, Q, is generally not known for complex systems such as molecular biofibril networks. To circumvent the lack of knowledge of the normalizing quotient, instead an estimate of $\langle U \rangle$ can be made based upon a series of K unique configurations of molecules, $\Gamma_j \{n_1,...,n_m\}$ for $j=1,\ldots,J$, such that

$$\langle U \rangle = \frac{1}{J} \sum_{j=1}^{J} U\left(\Gamma_j\right). \tag{19-4}$$

As K becomes large, the estimate of $\langle U \rangle$ approaches the expected minimum value of the internal energy of an ensemble of molecules. The simulation space must be initially seeded with molecules in a way that precludes infinite energy interactions (e.g. interactions that violate volume exclusion). After initially seeding of the simulation space, an initial energy is calculated based on the thermal properties of the system, the interaction potentials, and distances of the interaction sites. Next a molecule is chosen randomly and displaced a random distance generating a new configuration. The energy of the new configuration is calculated. As long as the quotient of thermodynamic probability of the new configuration is less than a random number ζ, generated on the interval of (0,1), the new configuration and associated energy are accepted. Otherwise, the molecule is returned to the starting position. The acceptance criteria is explicitly demonstrated by

$$\zeta \leq \exp\left[-\left(U'_j - U\left(j\right)\right)/kT\right], \tag{19-5}$$

where $U(j)$ denotes the baseline energy prior to reconfiguration and U'_j is the energy of the new system following the displacement of the randomly chosen molecule.

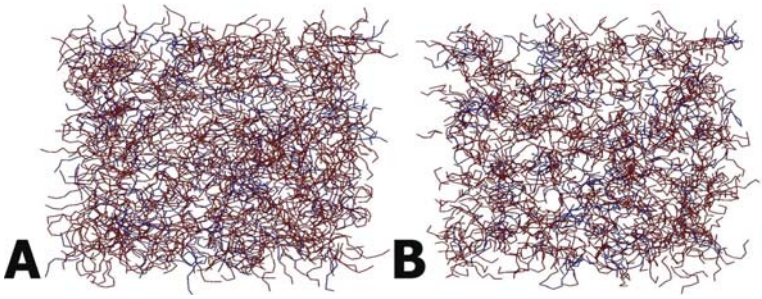


Figure 19-5. Collagen IV Network Generated with MH algorithm. The network is initialized as a collection of $\alpha_1\alpha_1\alpha_2$ (~80%) and $\alpha_3\alpha_4\alpha_5$ (~20%) monomers randomly selected in the simulation space. System energy is decreased as 7S domains are brought within bonding proximity of other 7S domains (~3 nm). Over the course of 5×10^6 Metropolis steps, the system begins to converge upon the global energy minimum of the system. (**a**) Initial network before energy minimization and (**b**) network following energy minimization via the Metropolis-Hastings algorithm. Notice the increase in heterogeneity

As the algorithm proceeds, the result is convergence to the minimal energy configuration of a network comprised of the initial fiber set. The algorithm is designed to prohibit convergence upon local energy minima and is sensitive to the level of correlation between random numbers. Consequently, a high-quality pseudo-random-number-generating algorithm is critical to ensure that results are statistically valid (see [153, 154] for a more thorough discussion on this topic). This application of the MH algorithm is thus a means to generate a network from the associated fundamental subunits using an energy minimization approach. An example of a collagen IV network generated using the MH algorithm is demonstrated in Figure 19-5.

19.4. GENERAL MODELING APPROACHES FOR BIOPOLYMER NETWORKS

19.4.1. Definitions

In the following, we define a biopolymer network to be a collection of interconnected fibers. Depending on the biopolymer of interest, a fiber may consist of a true fiber, a fibril, a filament, or a bundle of filaments, which themselves are composed of monomers, depending on the network of interest. Where two fibers interact, we define there to be a node. There are two major types of nodes: entanglements and cross-links. An entanglement is a point at which two or more fibers are in close proximity, such that the possibility of contact alters the deformation properties of the network. A cross-link is defined to be a point where two fibers are chemically linked together. A segment of length l_s is defined to be a piece of the fiber between two neighboring nodes.

We also define various types of polymer networks based on the flexibility of the fibers that make up the network. Flexibility is determined by the persistence length l_p of a fiber, which gives the typical length over which a fiber remains straight. For a fiber of length L parameterized by s, l_p is given by

$$\langle \cos [\theta(s) - \theta(0)] \rangle = \exp \left[-s/l_p \right], \tag{19-6}$$

where $\theta(s)$ is the tangent angle of the fiber with respect to its main axis [156] and $\langle x \rangle$ is the expected value of x. It can also be shown that $l_p = K_b / kT$, where K_b is the bending stiffness, k is the Boltzmann constant, and T is the temperature [156]. The bending stiffness is a function of the fiber's Young's modulus and the moment of inertia of the cross-sectional area. A flexible network, like rubber, is one where $l_p << l_s$. Such networks are dominated by the entropic stiffness of the segments [157]. For a semi-flexible polymer network, such as actin, $l_p \approx l_s$ [63, 64, 158]. These networks are considerably more complex because both mechanical and entropic properties of the fibers play a role in the network dynamics. On the other extreme are mechanical networks, such as collagen-I gels where $l_p >> l_s$, and the fibers have very long thermal persistence lengths and the entropic effects are negligible [84].

19.4.2. Affine Theory

A variety of approaches exist for modeling biopolymer networks. One common assumption employed is that the network deformation can be described as an affine transformation. An affine transformation preserves the collinearity of points and the ratio between distances. A typical example is that of simple shear, which maps the point (x,y) to $(x+\gamma y,y)$.

Affine theories have been used to describe the properties of flexible gels, such as rubber [157], where the persistence length is much shorter than the distance between nodes. Here, it has been found that the elastic modulus, G', scales with the cube of the mesh size, ξ. Mesh size is defined as the average of sphere diameter that fits inside the network without touching the fibers. For semiflexible biopolymer networks, the persistence length is on the same order of the mesh size, and the entanglement length l_e is used to describe the network. MacKintosh et al. [158] and Morse [63, 64] have developed an affine deformation theory for semiflexible biopolymer networks. By treating the polymer as an entropic worm-like-chain, they derive a force-displacement curve for an individual polymer chain to be:

$$F \sim \frac{K_b^2}{kT}\frac{dl}{l_e^4}, \tag{19-7}$$

where dl is the length change of a segment. The force-displacement curve is related to the modulus of the material by assuming there are ξ^2 fibers per area, and that $dl = \gamma l_e$, where γ is the shear strain. This gives

$$G' \sim \frac{K_b^2}{kT}\frac{dl}{\xi^2 l_e^3}. \tag{19-8}$$

To relate G' to the polymer concentration, one must first determine the dependency of K_b, ξ, and l_e on concentration. MacKintosh et al. [158], assume the fibers do not bundle and thus K_b is independent of c. Previous experiments [66] indicate that for non-cross-linked actin, $\xi \sim c^{-1/2}$. For l_e, various relationships have been used. For a densely cross-linked gel, $l_e \sim \xi$, and $G' \sim c^{2.5}$. However, the precise relationship between l_e and ξ may be more complex. On the other hand, modeling the chain as a fluctuating rod gives $l_e \sim \xi^{4/5}$ [158], whereas if one also assumes that the bending stiffness of the polymer depends on the polymer length, as is the case when polymer bending is dominated by shear (discussed below), then $l_e \sim \xi^{4/3}$. The affine theory has also been used to explain strain stiffening of biopolymer gels [70] as well as the negative normal stresses observed during shear [69]. Affine-deformation models have also been used to simulate the mechanical response of fibrillar tissues [159], including heart valve [160], cornea [161], skin [162], and articular cartilage [163], accounting for multiple co-existing networks or non-fibrillar tissue components as needed.

While the affine model predicts behavior in line with what has been observed experimentally, it is not clear that the assumption of affine deformation is valid at

the length scale of the fibers. The nonaffinity of a deformation can be measured in a number of ways [164] and is typically done by looking at the difference in length, angle, or vector difference between the observed deformation and that predicted for a purely affine deformation. Nonaffine deformations have been observed in practice [72, 102, 114, 117, 165, 166] but disagreement still exists on the applicability of the affine assumption for biopolymer networks [167, 168].

One difficulty with the affine assumption is that the network segments deform independently, and thus the details of network interactions are lost. Such an assumption allows a simpler material description that in some applications may be sufficient for the problem. Within this framework, however, there is no obvious way to account for fiber synthesis or degradation, nor does it allow one to model failure at the cross-links or in individual network fibers. The need for more detailed understanding of networks has led to the development of various non-afifne models, described in the next section.

19.4.3. Nonaffine Models

In modeling non-affine networks, there are three main choices the modeler must make: (1) the constitutive model for the individual fiber segments, (2) the properties of the nodal interactions, and (3) the network organization of the segments. The choice, in part, depends on what type of questions the modeler intends to answer. The individual segments may be treated as linear or nonlinear springs, which only stretch, or as beams or worm-like-chains, which also resist bending and torsion. Additional relationships may be needed to account properly for the bending stiffness if the segment is composed of a bundle of interacting filaments.

Nodes can be treated either as cross-links or entanglements. While macroscopic scaling theories account for both (discussed later), with the exception of Rodney et al. [169], all microstructural models presented here assume that fibers are chemically cross-linked, or sufficiently entangled that on the time scale of interest they are and unable to slip at the nodes. In addition, the analysis is greatly simplified by neglecting steric interactions between fibers, which may contribute to the mechanical response. Nodes may be treated as freely rotating pin joints, welded joints of fixed angle, or linear or torsional springs.

19.4.3.1. Spring Model

We begin by exploring networks of randomly oriented springs, studied by Kellomaki et al. [144]. In this model, each segment acts as a linear spring, and the springs are connected at freely rotating pin joints. Kellomaki et al. [144] showed that under small deformations, such a network is floppy and has zero shear modulus. That is, under small deformations, the network is able to rearrange itself without changing the length of any of the springs. The floppiness of the network can be explained in two ways. One is that for a network cluster to be rigid, it must be composed of triangles that share a common side. Such a structure requires the existence of points where three fibers overlap. In a randomly generated Mikado network, the probability

of three fibers intersecting at a single point is almost surely zero, so it is impossible for a stiff cluster of fibers to percolate the network. An alternative framework for analyzing network properties is based on Maxwell counting [170]. Consider a d dimensional space composed of N_v vertices connected by N_c segments. The condition number is defined to be the average number of segments that connect to a single node and is given by $z = 2N_c/N_v$. The total number of degrees of freedom in the network (ignoring rigid motions) is given by $N_f = N_{vd} - N_c$, where d is the spatial dimensionality. For a rigid network $N_f = 0$, giving $N_c = N_{vd}$, and requiring $z = 2d$, where d is the spatial dimensionalty. For a network of Mikado model structure, even if the free ends are removed (of condition number 1), we are left with vertices that are connected by 2–4 springs. Because the condition number is less than four the network is floppy. An important implication of this model is that the assumptions associated with the affine model discussed above are inconsistent. If a biopolymer gel is modeled as a network of randomly oriented springs, even if the springs are nonlinear, the network cannot resist shear at infinitesimal deformation. In contrast, springs in the affine model stretch immediately. Furthermore, the spring network does not deform affinely because it is able to rearrange itself under small deformations without changing the length of its springs.

Chandran and Barocas [73] have also studied random spring networks with the goal of modeling collagen gels. They studied networks generated by an artificial polymerization algorithm, described above, and these networks also have a condition number that is less than 4. Similar to Kellomaki et al. [144], they find that the network deformations are significantly different from affine and in particular, fibers are likely to reorient rather than stretch, thus leading to smaller stretch ratios than would be seen in the affine case even though the fiber orientation averaged over the entire population remained close to the affine value.

The Mikado and polymerization models are attractive in that the network architecture is reminiscent of biopolymer networks, but their main problem is that networks of zero modulus at small strains are unrealistic. To study rigid spring networks, Wyart et al. [171] explore the strain stiffening properties of networks formed by an alternative algorithm, in which the space is seeded with a number of nodes and then a condition number is imposed by connecting vertices that are close. Buxton and Clarke [172] have also studied beam networks formed in this way. This method allows one to explore the transition from floppy to rigid networks as the condition number increases. However, because this network architecture is not representative of most biopolymer networks, we do not discuss its properties further.

A fourth architecture for modeling biopolymers is the Arruda-Boyce eight-chain network model, used by Palmer and Boyce [173] for modeling actin networks. The model represents the network as a unit cell containing eight segments, each connecting a corner of the box to the center. Incompressibility is imposed on the cell such that even though the network alone is floppy, the network in combination with the incompressibility constraint is stiff. These models have the advantage of being easy to solve, but like the random networks of Wyart et al. above, they are not representative of true biopolymer networks. Nevertheless, by tuning the segment parameters, one can match experimental data for skin [174] and actin networks [173]. In modeling actin, Palmer and Boyce [173] based their force-displacement

curves on the theory of MacKintosh [158], described above. This modeling framework allows consideration of prestress, but unfortunately, gives no way to predict it in the network, so the prestress must be fit to each data set individually.

19.4.3.2. Beam Models

In light of the above the result that realistic, spring networks have $G' = 0$, it is necessary to account for the bending energy of the segments as well, or, at a minimum, to introduce torsional springs at the nodes [74]. Explicitly accounting for fiber bending is typically accomplished by treating each segment as a worm-like-chain (WLC) that contains both stretching and bending energy. Numerically, segment bending can be implemented either by using a discrete WLC model [175] or a finite element algorithm [145, 176], where the segments are represented as beams having both a stretching stiffness K_s, and a bending stiffness K_b. Often, the segments are treated as elastic rods using Euler-Bernoulli beam theory, and thus the mechanical stretching stiffness is given by $K_s = EA$ and $K_b = EI$, where E is the Young's modulus of an individual fiber, $A = \pi R^2$ is the cross-sectional area, R is the rod radius, and I is the area moment of inertia of the rod. The total energy in a filament is given by,

$$H = \frac{1}{2}K_b \int_0^L \left(\nabla^2 u\right)^2 dl + \frac{1}{2}K_s \int_0^L \varepsilon^2 dl, \tag{19-9}$$

where the segment is of length L, the transverse displacement is given by u, the curvature is given by $\nabla^2 u$, and the axial strain given by ε. From the above expressions, one can also define a spring stiffness for a segment. The mechanical stretching stiffness for a cylindrical fiber segment is given by $k_s = K_s/l_c$, while the bending stiffness is $k_b = K_b/l_c^3$, where l_c is the mean spacing between nodes.

If the segment consists only of a single isotropic, linear elastic filament of radius r, then $I = I_{\text{fil}} = \pi r^4/4$. However, in many biopolymer networks, including actin and collagen, the segment is in fact a bundle of fibers. If the bundles are tightly coupled together by a stiff and rigid cross-linker, as is the case for actin cross-linked by scurin [177], then a similar formula applies, $I = \pi R^4/4$, with R the radius of the bundle. In the case of loose intrasegment coupling, we instead have $I = N_{\text{fil}} I_{\text{fil}} = R^2/r^2 I_{\text{fil}} = \pi (Rr)^2/4$. In this case, the bundle is much more flexible as $I \sim R^2 r^2$ instead of R^4. There also exists a third, intermediate regime, determined by the nondimensional parameter $\alpha = k_x L^2/(EA\delta)$ [178, 179], in which k_x is the cross-link stiffness at a node, L is the segment length, and δ is the mean spacing between nodes. For small α, the coupling is weak and $K_b \sim E(Rr)^2$. For very large α, the coupling is strong and $K_b \sim ER^4$. For intermediate values of α, the formula for K_b is more complicated, depending upon L, δ, and k_x [178]. In the case of actin bundles, three regimes have been observed, depending upon the cross-linker used [179].

The first step in understanding the mechanical properties of these networks is to explore the G' behavior as a function of the network density. For the Mikado model,

we define two densities: a nondimensional density $q = NL^2/A$, and a dimensional density $\rho = NL/A$, where N is the number of fibers of length L (possibly containing multiple segments) in a box of area A, and ρ has units of [1/length]. In two dimensions, it is possible to link l_c directly to N, L, and A and at large q, it is given by $l_c = 2/q\pi$ [180]. The mechanical properties of these networks have been shown to depend critically on q [181], and also on the average number of nodes per segment L/l_c [143]. Here, we choose to describe the network mechanics in terms of L/l_c, though the two choices are equivalent [180]. For low L/l_c, the system is made up of isolated rods and small, unconnected clusters. In such a system, there is no connected path from one side to the other and $G' = 0$. At $L/l_c = 5.42$ [143], conductivity percolation occurs, meaning that a path exists connecting two opposite sides of the network. In the case that the nodes can resist rotation (e.g. welded joints), the system has also achieved rigidity percolation, and the connected component can resist deformation. In the case that the nodes are treated as freely rotating pin joints, rigidity percolation does not occur until $L/l_c = 5.93$ [143].

Above rigidity percolation, there are two mechanical regimes, based on whether the deformations are affine or nonaffine. In the case of $k_b << k_s$, the fibers are long and thin, and the spacing between cross-links is large. Since the bending stiffness is relatively low, the network responds to deformation by the bending of its fibers, which is inherently non-affine. For freely rotating cross-links, it has been shown both through simulation and through a self consistent analysis of the floppy modes of the system that $G' \sim k_b \left(L/l_c\right)^{3.67} \sim K_b \rho^{6.67} L^{3.67}$ [141], exhibiting an extreme sensitivity to the network density. The system behaves fundamentally differently when $k_s >> k_b$. In this regime, deformations are affine and $G' \sim k_s \sim EA\rho$, and the modulus depends only linearly on density. The critical length at which the transition occurs is given by $l_{\text{crti}} = L\left[\left(\rho - \rho_f\right) L\right]^{-2.84}$. Thus far, this scaling transition has not been fully explored in three-dimensional simulations. However, Huisman et al. [182] have studied artificially generated networks designed to be similar to actin, and Stein et al. [84] have studied collagen networks of realistic architecture. Both have found that at small deformations, the primary mode of energy storage is in bending, and at small strains the deformations are highly non-affine. This result lends further support to the idea that the affine assumption is erroneous for most biopolymer networks.

As discussed in more detail below, cross-linked biopolymer networks typically scale by $G' \sim c^{2-3}$, where c is the polymer concentration. This is quite different than either scaling regime for the Mikado model. Thus the importance of the above results is not in the specific scaling laws derived, but in the observation that there are two distinct mechanical regimes, one dominated by affine stretching and another dominated by nonaffine bending.

19.4.3.3. Entropic Beam Model

An additional level of detail that can be added to the model is the entropic component of the stretching stiffness of an individual filament. In this framework, the stretching stiffness k_s of a segment is modeled as two springs connected in series:

an elastic element $k_{\text{el}} = EA/l_c$, and an entropic element given by $k_{\text{en}} = K_b^2/kTl_c^4$. The total stretching stiffness is given by $k_s^{-1} = k_{\text{el}}^{-1} + k_{\text{en}}^{-1}$ and is dominated by the more compliant of the two elements. In the case that the entropic stiffness is weakest, we have $k_s \sim l_c^{-4}$, which is markedly different from $k_s \sim l_c^{-1}$ for purely mechanical networks. Fibrous networks exhibit G' scaling that is very sensitive to the cross-link behavior. For rigid cross-links, $G' \sim Ak_s + Bk_b$, where rather than acting in series, the stretching elements now act in parallel, with the stronger of the two dominating the elastic response. For flexible cross-links, however, in the case of inextensible fibers ($k_{\text{el}} \to \infty$), $k_b^{0.5}k_s^{0.5}$. For these systems too, it is found that there is a critical average segment length l_{crit}, such that for networks with $l_c < l_{\text{crit}}$, the deformations are affine and when $l_c > l_{\text{crit}}$, the deformations are nonaffine. Thus including the entropic properties of the network gives qualitatively different scaling laws for G', but the nonaffine/affine transition is still present.

19.4.4. Finite Strain

19.4.4.1. Strain Stiffening

The above section focused on the small-strain behavior of networks. Soft tissues, in particular, routinely deform beyond the small strain limit, ranging anywhere from 2% to over 40% strain depending on the tissue [76]. The typical load-deformation response of cells and soft tissues in uniaxial tension is non-linear, starting with a long, extensible toe region, followed by a linear and then exponential increase in the force. The stiffening observed at high strains is universal, but the cause of strain stiffening is unclear; the proposed mechanisms underlying it are dependent on the type of model used, and the mechanisms may be different for different biopolymers.

Storm et al. [70] have used the affine theory developed by MacKintosh [158] to show that the strain stiffening of a large number of polymers could qualitatively be explained by an affine deformation of a network of strain stiffening filaments. Similar behavior can also be produced by network reorganizations in which fibers are free to rotate with the deformation, both in spring [74, 149] and beam models [84, 146, 147, 182, 183]. It is likely that the precise nature of strain stiffening depends upon the specific properties of the biopolymers, as well as the manner in which they are organized. Tissues are denser and more cross-linked than gels, and their strain stiffening response may derive from both molecular entropic effects and ECM geometry. In some instances, the scale of a problem involving tissues may warrant the use of affine theories [166], although experiments show that at least some tissue fiber deformations are not affine [114, 166, 184].

19.4.5. Bridging Scales – Multiscale Behavior of Networks

19.4.5.1. Representative Volume Element

A common approach used in relating the macroscopic behavior of a material to its microstructure is to find a region in which the microstructure is structurally typical to the entire sample [185]. Such a region is referred to as a representative volume

element (RVE). RVEs possess a characteristic length scale that is at least an order of magnitude larger (and preferably larger) than the length scale of heterogeneity in the microstructure. As a result, the whole material can be subdivided into a repeating array of RVEs, joined at their boundaries. Because the RVE is periodic and it is similar mechanically to the whole material, analysis can be conducted on the RVE alone.

Once an RVE has been selected, the analysis that follows assumes that the microstructure deforms continuously with the macroscopic strain field (the affine assumption). Such is the approach taken with cellular solid models [139, 186], which have been used to study a variety of materials, including metals and plastics [139], bone [187–189], and connective tissue in the optic nerve head [190]. Cellular networks can be setup with idealized, regular geometries that permit analytical solutions, or they can be created with irregular structures and probed with the finite element method [141, 191]. Either way, the bulk properties of the material can be related to the microstructure.

The analysis, however, is not limited to the behavior of one archetypal RVE. The RVE mechanical response can vary spatially, as in homogenization theory, where RVEs in the material develop different levels of strain to accommodate the inhomogeneous macroscopic displacement field [192–194]. The strategies to link scales in soft tissues are more challenging because large deformations are possible; hence, techniques based on a small strain assumption, such as many forms of homogenization, will fail. More importantly, linking strategies that rely on periodicity cannot incorporate macroscopic heterogeneity.

Our group has developed a multi-scale computational model that relies on the method of volume averaging [195] to link the macroscopic level to the microscopic level [74, 117, 149]. Because a material average volume is formulated (i.e. a volume that deforms with the material) large deformations are easily addressed. Furthermore, macroscopic heterogeneity, manifested as regional differences in the local ECM microstructure, can be accommodated naturally by employing different RVE network structures, provided that the regional differences are larger than the scale of the RVE. To clarify, the RVE domain should be bigger than the scale of microscopic gradients but smaller than that of macroscopic gradients [196]. As a result, the model provides a means to study the dependency of macroscopic tissue mechanics on the underlying ECM microstructure, which for our purposes is typically represented as a network of collagen fibers contained within an RVE [74, 117, 149, 150]. Consequently, the remainder of this discussion applies to collagen fiber networks, but other networks (e.g. electrospun fibers [195]) can also be examined with the method provided their attributes are accounted for in the fiber constitutive equation and volume averaging equation detailed below.

19.4.5.2. Volume Averaging

In the model, the macroscopic domain is represented with a Galerkin finite element (FE) model (Figure 19-6). However, in place of a macroscopic constitutive equation, the stress needed for the FE solution is obtained by solving the force balance on the

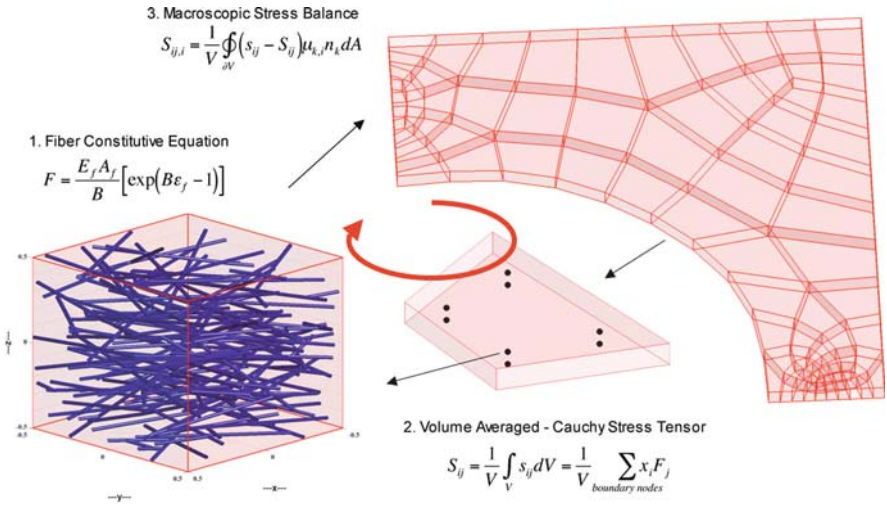


Figure 19-6. Multiscale modeling with volume averaging. The multiscale model relies on volume averaging theory to link scales. The macroscopic problem is represented using the Galerkin finite element method. RVEs containing fiber networks are centered at the integration points in the element, and the RVE boundaries are deformed with the macroscopic deformation field. Fiber forces in the network are volume averaged and the resulting macroscopic stress tensor is used in the macroscopic stress balance to solve for the new macroscopic displacement field. This process iterates going back and forth between scales until convergence is achieved

fiber network contained within an RVE. The RVEs are centered at the FE integration points, and their boundaries are displaced based on the macroscopic deformation field. Boundary displacements produce forces in the fibers that are transmitted via fiber crosslinks, with the result that the fibers in the network reorganize to achieve force equilibrium. The network fiber stress is averaged over the RVE to obtain the macroscopic average Cauchy stress tensor, which is then used in the macroscopic stress balance to determine the new macroscopic deformation field, and the process iterates until convergence is achieved.

The method utilizes three basic equations: (1) a constitutive equation to relate fiber stress to fiber strain (2) an equation that relates the average macroscopic stress to the volume average of the local fiber stresses (3) and an equation for the force balance at the macroscopic level. A fourth expression to incorporate rotational stiffness at the nodes can also be used [74]. A number of constitutive equations have been proposed for collagen fibers [160, 193, 198] that represent the fiber as strong in tension and weak in compression. In previous work [73, 149], we have found that the fiber constitutive equation used only has a minor influence on the macroscopic behavior. For convenience, we employ an exponential constitutive equation [160] to relate the fiber force, F, as

$$F = \frac{E_f A_f}{B}\left[\exp(B\varepsilon_f) - 1\right], \tag{19-10}$$

where E_f and B are constitutive constants, and A_f is the fiber cross-sectional area. The Green's strain of the fiber, ε_f, is given in terms of the fiber stretch ratio, λ_f, as $\varepsilon_f = 0.5\left(\lambda_f^2 - 1\right)$. Equation (19-10), at the low-strain limit, reduces to a linear model with elastic modulus E_f.

In volume averaging [195, 199], the macroscopic Cauchy stress tensor, S_{ij}, is determined by averaging the microscopic stress field, s_{ij}, over the RVE volume, V,

$$S_{ij} = \frac{1}{V}\int_V s_{ij} dV. \tag{19-11}$$

Here we use index notation with uppercase and lowercase letters to refer to macroscopic and microscopic variables, respectively. The microscopic stress can be rewritten as $s_{ij} = s_{kj}\delta_{ik}$, where δ_{ik} is the Kronecker delta. Because the gradient of the direction vector, $\boldsymbol{x}$, is equivalent to δ_{ik} ($\nabla x = x_{i,j} = \delta_{ij}$), Eq. (19-11) can be rewritten as

$$S_{ij} = \frac{1}{V}\int_V s_{kj} x_{i,k} dV = \frac{1}{V}\int_V \left(s_{kj} x_i\right)_{,k} dV - \frac{1}{V}\int_V s_{kj,k} x_i dV. \tag{19-12}$$

The second term on the RHS vanishes because microscopic equilibrium requires that $s_{kj,k} = 0$. Applying the divergence theorem allows the macroscopic stress to be calculated as integral of the RVE surface tractions, t_j, over the RVE surface,

$$S_{ij} = \frac{1}{V}\oint_{\partial V} n_k s_{kj} x_i dA = \frac{1}{V}\oint_{\partial V} x_i t_j dA. \tag{19-13}$$

The tractions occur at the locations where network segments intersect the RVE boundary (cross-links). For thin segments, x varies little over the segment-boundary intersection. Thus, $\int n_k s_{kj} x_i dA \approx x_i \int n_k s_{kj} dA = x_i F_j$, and the components of S_{ij} are given in terms of the crosslink positions, x, and forces, F, as

$$S_{ij} = \frac{1}{V}\sum_{\text{boundary nodes}} x_i F_j, \tag{19-14}$$

The final equation needed is the macroscopic stress balance. Since the averaging volume is material and changes with the macroscopic displacement, additional terms must be incorporated (see Chandran [74]). The advantages of material description are (1) it is consistent with how the microstructure deforms, (2) it satisfies the mass balance implicitly, and (3) the macroscopic gradients are naturally applied as the boundary conditions of the RVE [74].

The macroscopic stress balance is given as

$$S_{ij,i} = \frac{1}{V} \oint_{\partial V} \left(s_{ij} - S_{ij} \right) u_{k,i} n_k dA, \tag{19-15}$$

where u is RVE boundary displacement and n is the unit normal vector. The right hand side of Eq. (19-8) arises from the correlation between inhomogeneous displacement of the RVE boundary and local inhomogeneities in the stress field. In the case of a fixed RVE, the RHS would be zero.

19.5. APPLICATIONS TO BIOPOLYMERS

Now that the general methods used to model biopolymer networks have been discussed, we examine the application of network models to specific problems.

19.5.1. Actin

Actin is a popular choice for microstructural analysis due to its critical role in a number of cellular events and biological processes, including cell motility [179]. Of particular interest is the wide spectrum of actin cross-linkers, whose effect on network formation and mechanics has important implications for normal cell function. When no cross-linker is present, the networks are extremely compliant ($G' < 0.5$ Pa) and elasticity scales with concentration as $G' \sim c_a^{7/5}$ [200]. The addition of a cross-linker can bring the modulus to 100 Pa or larger, clearly demonstrating their importance in network formation. Cross-linkers, however, can serve two functions. First, they can group individual filaments into a larger bundle, which can strongly influence the bending stiffness [179, 201]. Second, they can connect filaments and bundles together to form a network. Cross-linkers vary in length with shorter molecules, such as scurin and fascin, forming relatively tight bundles whereas longer molecules, such as filamin and α-actinin form looser bundles. Heavy meromyosin cross-links while forming no bundles at all. The effect that various cross-linkers have on the actin networks is summarized in Table 19-2. Remarkably, scurin [60, 68], fascin [202], and HMM [203] all have similar effects in terms of the scaling of G' with respect to the actin concentration and cross-linker ratio, with

Table 19-2. Effect of cross-linkers on actin network properties

Cross-linker	None	Filamin [67]	HMM [201]	Fascin [61]	Scurin [60, 68]
Bundle formation	None	Loose	None	Tight	Tight
G' (Pa)	0.1–0.5	1–10	0.1–100	0.1–100	0.1–100
$G' \sim {c_A}^x$	1.4		2.2	2.4	2.5
$G' \sim R^x$ (large R)	N/A	0.4	1.2	1.5	2
$G' \sim {\gamma_{\text{crit}}}^x$	N/A		–1.0	–0.4	–0.6

$G' \sim c_a^{(2.35\pm0.15)} R^{(1.6\pm0.4)}$. An additional parameter that is tracked is the critical strain $\gamma_{\text{crit}} \sim R^{(-0.7\pm0.3)}$, which indicates the onset of strain stiffening, and again is relatively similar for the three different cross-linkers.

Many models exist that have been used to explain some of these data, including the affine stretching model of worm-like chains [68, 158], the nonaffine 8-chain model [173], and the nonaffine bending model [167, 202]. All three models have also been able to explain the strain stiffening behavior of biopolymers [70, 147, 173]. In vivo, turnover of the actin network may contribute to its apparent viscosity. That is, a stressed fiber may disassemble and be replaced by new, unstressed fibers. The significance of this phenomenon varies with cell type, phenotype, and activity. While non-affinity has been directly observed at short length scales in scurin-cross-linked actin networks [102], the community does not yet agree upon whether such nonaffinity is sufficient to invalidate the affine theory [168]. The 8-chain model of Palmer and Boyce [173] requires one to refit the network prestress at each actin density. Thus their model makes the prediction that lower density networks have higher degrees of prestress, but such a prediction has yet to be validated.

Finally, it has been observed that the maximum strain that a gel can withstand decreases with increasing density [204]. This is hypothesized to be due to a shortening of the space between cross-links, which according to entropic stiffening hypothesis, means that the fibers reach their maximum state of strain sooner [158].

19.5.2. Microtubules, IFs, and the Cytoskeleton

Microtubules and IFs have been cast into gels and subjected to rheology tests to determine their individual mechanical characteristics [205, 206]. However, the networks were formed in vitro from purified monomer and may differ substantially from those formed inside a cell. It is important to understand the individual properties of these proteins, but how they integrate with actin to form the cytoskeleton is the ultimate goal, and much remains to be learned.

Wang and Stamenovic explore the contribution of IFs to cellular mechanics by measuring cell stiffness to applied stress in adherent wild-type and vimentin-deficient fibroblasts through magnetic twisting cytometry [207]. At high applied stress (>> 10 dynes/cm^2), the stiffness of the vimentin-deficient fibroblasts is much smaller than the wild-type fibroblasts, while at a stress of 10 dynes/cm^2, the stiffness is comparable. A six-strut tensegrity model (discussed below) was able to replicate the stiffening that resulted from cytoskeletal fiber realignment.

Microtubules and IFs are integrated into the cytoskeleton, and therefore can affect the properties of the whole cell. One perspective on the structure-function relationship between the cell, its cytoskeleton, and the extracellular matrix is the hypothesis that the cell is a tensegrity (tensional integrity) structure [26, 208, 209]. In this model, the stability of the cytoskeleton is derived from a balance between a continuous filament network (actin and intermediate filaments) under tension and isolated compression-resistant elements (microtubules and thick actin bundles). Without internal tension, or "prestress", which can be generated through the cell's contractile machinery, the cytoskeleton would collapse. External forces, which are

transmitted from the ECM to the cytoskeleton through focal adhesions, cause the cytoskeleton to reorganize and stabilize until equilibrium is achieved. Support for this view appears to be based largely on its intuitive appeal and model predictions that match cell stiffening behavior with increasing tension and surface attachments [210, 211]. The non-linear behavior observed derives from geometric changes in the network, a behavior also observable in random fiber networks without compression elements. Although the concept of tensegrity is attractive, and experimental evidence shows that microtubules can buckle under cellular loading [212], the exact nature of the complex interactions that define cytoskeletal mechanics remains poorly understood.

19.5.3. Spectrin

Spectrin has been studied as triangulated networks of Hookean springs of non-zero force-free length and finite maximum length [212]. Triangulated networks are generated from infinitely thin hard rods with six-fold vertices. Under compression, the six-fold symmetric structures undergo a phase transition to two-fold network symmetry as studied analytically at zero temperature or through Monte Carlo simulation with a non-zero temperature [212]. In subsequent studies, a six-fold symmetric network of polymer chains, representing the actin-spectrin cytoskeleton of the erythrocyte, is generated and the geometrical and elastic properties are determined [214] and found to be in agreement with the shear modulus, of 6.6×10^{-3} dynes/cm at 25°C, for the erythrocyte cytoskeleton as determined from micromechanical techniques [215].

With respect to modeling the macroscopic erythrocyte structure, an intertriangulated network of chains becomes unwieldy. Consequently, Boey, et al., represented the spectrin chains using a worm-like chain potential providing a tractable and more physically realistic representation of an intertriangulated spectrin network than the original Hookean spring representation [214]. An ensemble-averaging technique was applied to non-axisymmetrical deformed shapes, analogous to an erythrocyte undergoing micropipette aspiration, demonstrating how the triangulated mesh of the spectrin-actin cytoskeleton imposes the macroscopic geometry of the erythrocyte [216]. Discher and colleagues' simulation results from three structural models of the spectrin network attached to a bilayer suggest that the network exists in a prestressed condition of compression balanced through tension created by the lipid bilayer [216]. Lee et al. later confirmed that the prestressed erythryocyte membrane is capable of sustaining large anisotropic strains using fluorescently-patterned photobleaching of a rhodamine phalloidin-labeled spectrin-actin cytoskeleton [214, 217].

Additionally, the equilibrium shape of the human erythrocyte has been investigated using spectrin-level energetics [218]. Li and colleagues populated spherical and biconcave structures with spectrin networks capable of 2,3,. . .,9 element junctions. After the initial shape was populated, cytosol was removed allowing the shape to deflate with fixed spectrin connectivities. Coarse-grained molecular dynamics

was employed to find the equilibrium shape of the deflated RBC employing a worm-like chain free energy model for the spectrin tetramer links. As a consequence of the hypothesis that spectrin networks are constantly undergoing remodeling at some sufficiently small characteristic time scale, Li *et al.*, employed a liquefied network structure evolution algorithm to relax the in-plane shear elastic energy of the macroscopic network shape which permits the evolution to discocyte and stomacyte shapes based on the approach of Discher and colleagues [213, 214, 216, 217].

19.5.4. Collagen I

Similarly to actin, collagen I has been modeled extensively due to its abundance and central structural role in many tissues. For this review, we focus on network models, where the fibers are modeled as tension resisting springs [73] or tension and moment resisting beams [84]. As with the networks applied to actin, these models show that significant matrix restructuring occurs in a highly non-affine manner. In addition, uniaxial deformations applied to random networks result in nonlinear stiffening, which arises from fiber rotations that gradually lead to fiber stretch. In the small-strain limit, most of the energy in the network is stored in fiber and cross-link bending (when bending is accounted for). At large strains the primary energy storage mode is fiber stretching, and thus the large strain modulus scales linearly with density [84].

Our multiscale model was originally developed for collagen gels, which appear smooth and continuous on the macroscopic scale but are in fact composed of discrete fibers. The model was applied to collagen gels [74, 149] and extended to complex geometries [219] by means of a sophisticated computational environment. A representative example can be seen in Figure 19-7, which shows the stretching of a

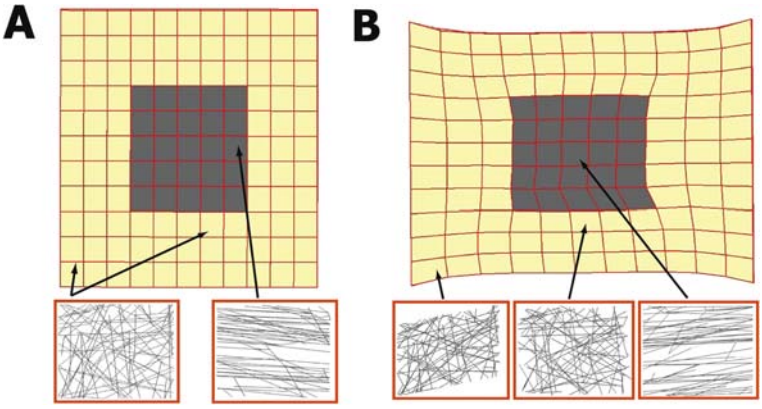


Figure 19-7. Heterogeneous test sample. A model system was constructed with a highly aligned central "wound" region and a more isotropic surrounding region. The sample was stretched uniaxially to 30% strain. The highly aligned central region deformed less than the surrounding isotropic region. Images are 2-D projections of a 3-D result

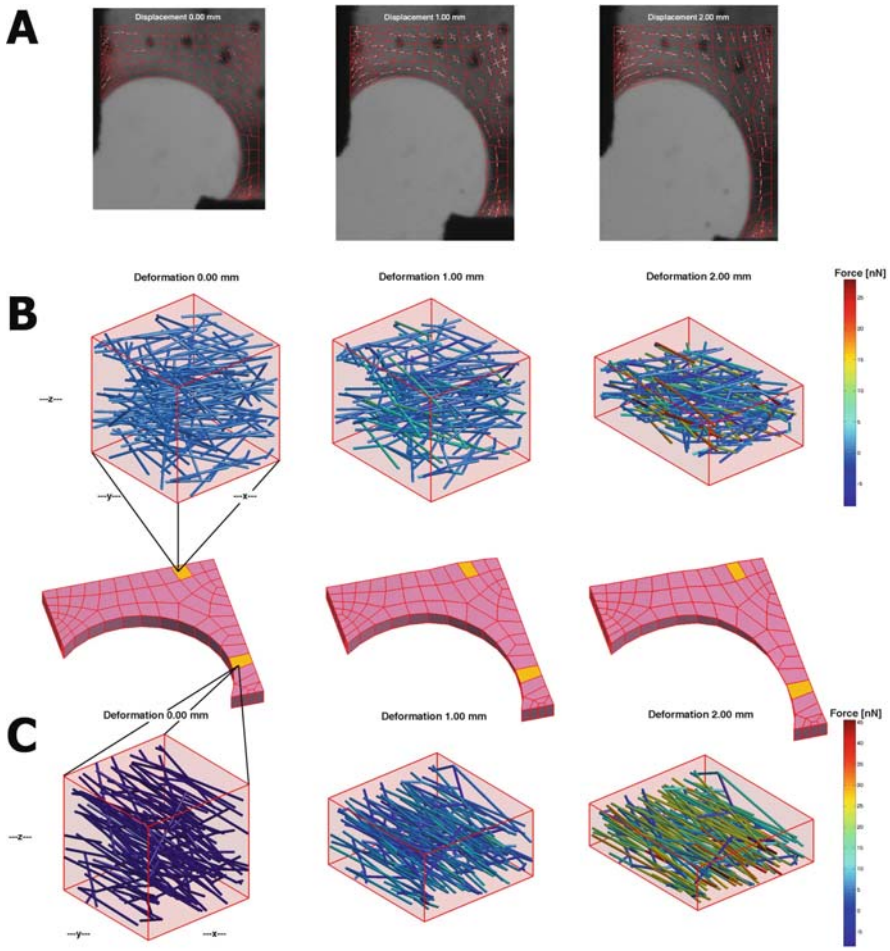


Figure 19-8. Image based multiscale modeling. One quadrant of a cell-seeded, compacted collagen gel was tested mechanically while microstructural orientation and strength of alignment was measured using polarized light. A unique 3-D, interconnected fiber network was generated for each finite element to match the average experimental values in the corresponding location in the experiment. (**a**) The model predictions (*red*) are overlaid on the experiment in white. For an off-axis hold test, in which the horizontal axis remained stationary while the vertical axis was displaced to a stretch ratio of 1.3, the model reasonably captured the local kinematics with some obvious differences. (**b**) The networks underwent significant restructuring in a location dependent manner. The network near the top right corner started off with moderate alignment in the horizontal direction. Fibers were free to rotate towards the vertical and stretch to accommodate the strain created by movement of the vertical grip. (**c**) This network, which was already aligned in the vertical direction increased in alignment with the stretch

sample with inhomogeneous structure. The model was also applied to electrospun synthetic polymer networks [197], demonstrating its generality.

As noted earlier, an important challenge in such modeling is acquiring structural information, and as the sophistication of the model increases, so too does its need

for structural specification. Sander et al. [117] have recently shown that the multiscale scheme can be specified based on polarized-light imaging, with each finite element in the model set at the average orientation measured for it in a sample. The sample was subsequently stretched in an off-axis hold test, which was concurrently simulated. The force and strain results were compared between the model and experiment, and agreement was good with fitting parameters based only on the constitutive equation for the fibers (an example of this kind of simulation is presented in Figure 19-8).

Because many tissues contain multiple components, a rule-of-mixtures model was applied to arterial wall mechanics [150], with a network phase representing the collagen fibrils in the vessel wall and an incompressible continuous phase representing the non-collagenous tissue (Figure 19-9 shows typical simulation results and comparison to Experiment). The model agreed well with experimental data on decellularized porcine carotid artery [220] and contained an important result: it was predicted that the non-collagenous material is in compression even when the vessel is inflated. That is, the role of the non-collagenous material in the model is to prevent vessel wall and/or lumen collapse, not solely to provide elastic recoil.

In some tissues, it has been observed that collagen fibrils/fibers/bundles are undulated and possess "crimp", and it has been hypothesized that their nonlinear stress-strain response is in part the result of collagen fibers straightening out [221]. When undulations were incorporated into a network model, the effect was to delay the stiffening response [147].

Affine theories for collagen have also fit well to some experimental data when fiber-level properties are not required to match fiber level measurements [160, 222, 223].

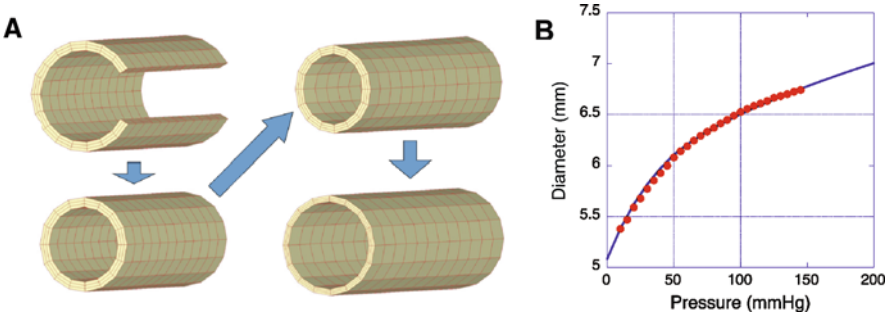


Figure 19-9. Simulated inflation of a decellularized artery [150]. (**a**) The simulation involves four steps: an initial open artery is constructed, it is closed to produce a prestressed artery, the artery is stretched longitudinally, and finally it is inflated. (**b**) Pressure-diameter curve for the simulation (*line*) and experiment (*dots*, [220]) are in good agreement with only two fitting parameters, the fiber stiffness and the matrix stiffness

19.5.5. Type IV Collagen

In spite of the importance of basement membrane, structural modeling has lagged behind that of collagen I or actin. A notable exception is the recent work of Burd [224]. Burd examined two phenomenological models (a linear elastic model and a Fung-type hyperelastic model), concluding neither model correlates satisfactorily with the mechanical properties of the collagen IV network present in the human lens capsule. Instead, a regular hexagonal lattice of pin-jointed bars was created and subsequently perturbed randomly to create an irregular polygonal structure, as used by Cavalcante et al. [225] for collagen-elastin networks. Representing the lens capsule as an irregular lattice of collagen IV embedded in a hyperelastic sheet correlated well with published data.

19.5.6. Fibronectin, Laminin, and the ECM

DiMilla and colleagues studied cell migration through a mathematical model elucidating the dependence of migration velocity on cellular mechanics and surface receptors binding with complementary ligands present in the extracellular substratum such as fibronectin and laminin [226]. They conclude that cytoskeletal force generation, cellular polarity, and adhesion dynamics are required for persistent cell motility. The model predicts how cell speed varies with a variety of phenomena including cytoplasmic rheology, intracellular contractile force, receptor/ligand densities and kinetics [226]. Ingber and Wang explored how the mechanical tension and the extracellular matrix through adhesion to the ligand fibronectin influence cytoskeletal mechanics [227]. Magnetic beads were coated with fibronectin of varying density, and a mechanical tension was applied to adherent endothelial cells through a magnetic twisting device. Ingber and Wang find that the cytoskeletal stiffness increases proportionally with the applied stress controlled by the magnetic twisting device and fibronectin density. A model is presented coupling cell mechanics to the applied stress [227]. Bischofs et al. developed a tension-elasticity model to correlate cell morphology as a consequence of adhesion to fibronectin-patterned substrates [227]. Bischofs derived a modified Laplace law from analytical results and computer simulation to describe filamentous network mechanics and contractility [227]. Such results demonstrate the coupling of extracellular adhesion influences cytoskeletal organization and cell shape.

19.6. SUMMARY

Two major challenges, in our opinion, remain for the community. The first is describing the segments and nodes in the network. Much work is now being done at the molecular scale [229], but we do not yet have network-level models that can incorporate molecular-level information. A similar rise in single-molecule and single-fiber experimental methods [19, 20, 41, 230] has not yet resulted in better fiber models in networks, which are still largely restricted to simple descriptions (springs, beams, or worm-like chains, none of which truly captures the full range

of behaviors seen at the molecular/fiber level). The challenge is even greater when nodes are considered. Pin joints, rigid crosses, and moment springs are all idealizations that do not capture the complex molecular interactions that characterize the junction between two fibers, and they are assuredly different for different biopolymers. It is expected that future scientists, armed with improved computer hardware and software, will incorporate our new knowledge of molecular detail as it arises.

The other major challenge, as suggested by our choice of title, is understanding how a network of many components can be interpreted in terms the one continuous material it appears to be on the macroscopic scale. We have seen in this paper how numerous researchers have explored the relationship between fibers and how the network properties vary with, for example, fiber or cross-link density, but a vast majority of the work has focused on these phenomena from a physical or material standpoint, working on the assumption of macroscopic homogeneity and often emphasizing a single component. While these studies represent a critical first step in any analysis, we need methods to incorporate multiple components and, as often occurs in biological systems, heterogeneity at all length scales. Only when these phenomena are understood will we truly be able to understand the behavior of the one tissue in terms of its many components.

19.7. NOMENCLATURE

A	Cross-sectional area of a fiber, [m^2]
B	Constitutive constant relating to fiber force
c	Concentration of polymer, [mol/L]
d	Spatial dimension
dl	Differential change in fiber length, [m]
E	Young's modulus of an individual fiber, [Pa]
$U(j)$	Exact energy of an ensemble of molecules in configuration j, [J]
E_f	Fiber constitutive constant, [N]
U'_j	Exact energy of an ensemble of molecules in configuration j with a slight perturbation to one randomly chosen molecule, [J]
$\langle U \rangle$	Expected average energy of an ensemble of molecules, [J]
F	Force along a fiber, [N]
$G'(\omega)$	Elastic modulus, [Pa]
$G''(\omega)$	Loss modulus, [Pa]
H	Total energy of a filament, [J]
I	Area moment of inertia of a rod, [m^4]
I_{fil}	Area moment of inertia of a filament, [m^4]
J	The last configuration for an ensemble of molecules undergoing an energy minimization scheme
K_b	Bending stiffness of a material, [N•m^2]
K_s	Stretching stiffness of a material, [N]
j	A series of configurations for an ensemble of molecules
k	Boltzmann constant, $1.3806503 \times 10^{-23}$ [J/K]

k_b	Bending stiffness of an individual fiber segment of length l_s, [N/m]
k_{el}	Elastic component of stretching stiffness of an individual fiber of length l_s, [N/m]
k_{en}	Entropic component of stretching stiffness of an individual fiber of length l_s, [N/m]
k_s	Stretching stiffness of an individual fiber in a network, [N]
k_x	Cross-link stiffness at a node, [N]
L	Length of a segment in a fiber network, [m]
l_c	Mean displacement between nodes in a network, [m]
l_{crit}	Critical length between cross-links through which a transition between affine and non-affine deformations, [m]
l_e	Entanglement length for a fiber, [m]
l_p	Persistence length for a fiber, [m]
l_s	Length of a fiber segment [m]
m	General representation to denote the last molecule in an ensemble of molecules
N	Number of fibers in a network
N_c	Number of connecting segments for N vertices in a d dimensional space
N_f	Number of network degrees of freedom in a network
N_{fil}	Number of filaments in a network
N_v	Number of nodes in a network
n_i	Denotes the ith molecule of an ensemble of m molecules
n_k	Normal vector
Q	Normalization quotient for calculating the average energy of an ensemble of molecules
q	A non-dimensional network density
R	Ratio of cross-linking component to actin concentration in a solution, or the radius of an individual fiber, [m]
r	Radius of an elastic filament, [m]
S_{ij}	Cauchy stress tensor, [Pa]
s	Axial parameterization of fiber length, [m]
s_{ij}	Microscopic stress field, [Pa]
T	Absolute temperature, [K]
t	Time, [sec]
t_j	Surface traction, [Pa]
u	Transverse displacement, [m]
$u_{k,i}$	Boundary displacement, [m]
$\nabla^2 u$	Curvature, [m^{-1}]
V	Volume, [m^3]
x	Power law exponent relating biopolymer concentration to elastic modulus
y	Power law exponent relating cross-linker-polymer ratio and elastic modulus

z	Condition number of a network (average number of fibers that connect to a node)
α	Dimensionless parameter
δ	Mean spacing between nodes in a fiber network, [m]
δ_{ik}	Kronecker delta
ε	Axial strain
ε_f	Green's strain of a fiber
$\gamma(\omega t)$	Shear strain
$\dot{\gamma}(\omega t)$	Shear strain rate, [s^{-1}]
$\Gamma_k \{n_1,...,n_m\}$	An individual configuration of m molecules
$\sigma(\omega,t)$	Stress, [Pa]
λ_f	Stretch ratio of a fiber
θ	Angular parameterization of the tangent angle of a fiber with respect to the longitudinal fiber axis, [radians]
ρ	Total fiber length per unit area of a 2-D Mikado network, [m^{-1}]
ρ_c	Critical fiber length per unit area of a 2-D Mikado network at which percolation occurs, [m^{-1}]
ξ	A random number on the interval (0,1), or the mesh size of a network
ω	Frequency, [radians/s]

REFERENCES

1. V. Vogel and G. Baneyx, "The tissue engineering puzzle: a molecular perspective," *Annu. Rev. Biomed. Eng.*, vol. 5, pp. 441–463, 2003.
2. B. I. Shraiman, "Mechanical feedback as a possible regulator of tissue growth," *Proc. Natl. Acad. Sci. USA*, vol. 102, pp. 3318–3323, Mar 1, 2005.
3. Z. H. Syedain, J. S. Weinberg and R. T. Tranquillo, "Cyclic distension of fibrin-based tissue constructs: evidence of adaptation during growth of engineered connective tissue," *Proc. Natl. Acad. Sci. USA*, vol. 105, pp. 6537–6542, May 6, 2008.
4. F. Grinnell, "Fibroblast biology in three-dimensional collagen matrices," *Trends Cell Biol.*, vol. 13, pp. 264–269, May, 2003.
5. F. Grinnell and C. H. Ho, "Transforming growth factor beta stimulates fibroblast-collagen matrix contraction by different mechanisms in mechanically loaded and unloaded matrices," *Exp. Cell Res.*, vol. 273, pp. 248–255, Feb 15, 2002.
6. M. Eastwood, V. C. Mudera, D. A. McGrouther and R. A. Brown, "Effect of precise mechanical loading on fibroblast populated collagen lattices: morphological changes," *Cell Motil. Cytoskeleton*, vol. 40, pp. 13–21, 1998.
7. S. Nakagawa, P. Pawelek and F. Grinnell, "Long-term culture of fibroblasts in contracted collagen gels: effects on cell growth and biosynthetic activity," *J. Invest. Dermatol.*, vol. 93, pp. 792–798, Dec, 1989.
8. F. Grinnell, M. Zhu, M. A. Carlson and J. M. Abrams, "Release of mechanical tension triggers apoptosis of human fibroblasts in a model of regressing granulation tissue," *Exp. Cell Res.*, vol. 248, pp. 608–619, May 1, 1999.
9. M. Kjaer, "Role of extracellular matrix in adaptation of tendon and skeletal muscle to mechanical loading," *Physiol. Rev.*, vol. 84, pp. 649–698, Apr, 2004.

10. C. F. Burgoyne, J. C. Downs, A. J. Bellezza, J. K. Suh and R. T. Hart, "The optic nerve head as a biomechanical structure: a new paradigm for understanding the role of IOP-related stress and strain in the pathophysiology of glaucomatous optic nerve head damage," *Prog. Retin. Eye Res.*, vol. 24, pp. 39–73, 2005.
11. R. L. Gleason and J. D. Humphrey, "A mixture model of arterial growth and remodeling in hypertension: altered muscle tone and tissue turnover," *J. Vasc. Res.*, vol. 41, pp. 352–363, Jul–Aug, 2004.
12. J. Khoshnoodi, V. Pedchenko and B. G. Hudson, "Mammalian collagen IV," *Microsc. Res. Tech.*, vol. 71, pp. 357–370, May, 2008.
13. C. Frieden, "Polymerization of actin: mechanism of the Mg^{2+}-induced process at pH 8 and 20 degrees C," *Proc. Natl. Acad. Sci. USA*, vol. 80, pp. 6513–6517, Nov, 1983.
14. C. Frieden, "Actin and tubulin polymerization: the use of kinetic methods to determine mechanism," *Annu. Rev. Biophys. Biophys. Chem.*, vol. 14, pp. 189–210, 1985.
15. C. Neidl and J. Engel, "Exchange of ADP, ATP and 1: N6-ethenoadenosine 5′-triphosphate at G-actin. Equilibrium and kinetics," *Eur. J. Biochem.*, vol. 101, pp. 163–169, Nov 1, 1979.
16. A. Wegner, "Spontaneous fragmentation of actin filaments in physiological conditions," *Nature*, vol. 296, pp. 266–267, Mar 18, 1982.
17. A. Wegner and P. Savko, "Fragmentation of actin filaments," *Biochemistry*, vol. 21, pp. 1909–1913, Apr 13, 1982.
18. A. Patkowski, W. Eimer, J. Seils, G. Schneider, B. M. Jackusch and T. Dorfmueller, "The molecular dimensions of G-actin in solution as studied by dynamic light scattering," *Biopolymers*, vol. 30, pp. 1281–1287, 1990.
19. H. Kojima, A. Ishijima and T. Yanagida, "Direct measurement of stiffness of single actin filaments with and without tropomyosin by in vitro nanomanipulation." *Proc. Natl. Acad. Sci. USA*, vol. 91, pp. 12962, 1994.
20. A. Ott, M. Magnasco, A. Simon and A. Libchaber, "Measurement of the persistence length of polymerized actin using fluorescence microscopy," *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics*, vol. 48, pp. R1642–R1645, Sep, 1993.
21. W. Kabsch and J. Vandekerckhove, "Structure and function of actin," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 21, pp. 49–76, 1992.
22. H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore and J. E. Darnell, *Molecular Cell Biology*, 4th ed. New York: W.H. Freeman & Company, 1999.
23. D. Boal, *Mechanics of the Cell*. Cambridge: Cambridge University Press, 2002, 406pp.
24. P. Cossart, "Actin-based motility of pathogens: the Arp2/3 complex is a central player," *Cell. Microbiol.*, vol. 2, pp. 195–205, Jun, 2000.
25. M. K. Gardner and D. J. Odde, "Modeling of chromosome motility during mitosis," *Curr. Opin. Cell Biol.*, vol. 18, pp. 639–647, Dec, 2006.
26. D. E. Ingber, "Tensegrity: the architectural basis of cellular mechanotransduction," *Annu. Rev. Physiol.*, vol. 59, pp. 575–599, 1997.
27. D. Stamenovic and N. Wang, "Invited review: engineering approaches to cytoskeletal mechanics," *J. Appl. Physiol.*, vol. 89, pp. 2085–2090, Nov, 2000.
28. H. Herrmann, H. Bar, L. Kreplak, S. V. Strelkov and U. Aebi, "Intermediate filaments: from cell architecture to nanomechanics," *Nat. Rev. Mol. Cell Biol.*, vol. 8, pp. 562–573, Jul, 2007.
29. D. E. Discher, R. Winardi, P. O. Schischmanoff, M. Parra, J. G. Conboy and N. Mohandas, "Mechanochemistry of protein 4.1's spectrin-actin-binding domain: ternary complex interactions, membrane binding, network integration, structural strengthening," *J. Cell Biol.*, vol. 130, pp. 897–907, Aug, 1995.
30. C. Picart, P. Dalhaimer and D. E. Discher, "Actin protofilament orientation in deformation of the erythrocyte membrane skeleton," *Biophys. J.*, vol. 79, pp. 2987–3000, Dec,2000.

31. D. E. Discher and N. Mohandas, "Kinematics of red cell aspiration by fluorescence-imaged microdeformation," *Biophys. J.*, vol. 71, pp. 1680–1694, Oct, 1996.
32. K. E. Kadler, "Extracelluar matrix 1: fibril-forming collagens" *Protein Prof.*, vol. 2, pp. 491–619, 1995.
33. P. Fratzl, *Collagen: Structure and Mechanics*. New York: Springer, 2008, 506 pp.
34. L. C. Junqueira, J. Carneiro and R. O. Kelley, *Basic Histology*, 9th ed. New York: McGraw-Hill, 1998.
35. F. H. Silver, J. W. Freeman and G. P. Seehra, "Collagen self-assembly and the development of tendon mechanical properties," *J. Biomech.*, vol. 36, pp. 1529–1553, Oct, 2003.
36. A. O. Brightman, B. P. Rajwa, J. E. Sturgis, M. E. McCallister, J. P. Robinson and S. L. Voytik-Harbin, "Time-lapse confocal reflection microscopy of collagen fibrillogenesis and extracellular matrix assembly in vitro," *Biopolymers*, vol. 54, pp. 222–234, Sep, 2000.
37. K. E. Kadler, D. F. Holmes, J. A. Trotter and J. A. Chapman, "Collagen fibril formation," *Biochem. J.*, vol. 316 (Pt 1), pp. 1–11, May 15, 1996.
38. Y. L. Sun, Z. P. Luo, A. Fertala and K. N. An, "Direct quantification of the flexibility of type I collagen monomer," *Biochem. Biophys. Res. Commun.*, vol. 295, pp. 382–386, Jul 12, 2002.
39. J. S. Graham, A. N. Vomund, C. L. Phillips and M. Grandbois, "Structural changes in human type I collagen fibrils investigated by force spectroscopy," *Exp. Cell Res.*, vol. 299, pp. 335–342, Oct 1, 2004.
40. H. Miyazaki and K. Hayashi, "Tensile tests of collagen fibers obtained from the rabbit patellar tendon," *Biomed. Microdev.*, vol. 2, pp. 151–157, 1999.
41. J. A. van der Rijt, K. O. van der Werf, M. L. Bennink, P. J. Dijkstra and J. Feijen, "Micromechanical testing of individual collagen fibrils," *Macromol. Biosci.*, vol. 6, pp. 697–702, Sep 15, 2006.
42. Z. L. Shen, M. R. Dodge, H. Kahn, R. Ballarini and S. J. Eppell, "Stress-strain experiments on individual collagen fibrils," *Biophys. J.*, vol. 95, pp. 3956–3963, 2008.
43. L. Yang, K. O. van der Werf, C. F. Fitie, M. L. Bennink, P. J. Dijkstra and J. Feijen, "Mechanical properties of native and cross-linked type I collagen fibrils," *Biophys. J.*, vol. 94, pp. 2204–2211, Mar 15, 2008.
44. D. J. S. Hulmes, "Collagen diversity, synthesis and assembly," in *Collagen: Structure and Mechanics*, P. Fratzl, Ed. New York: Springer, 2008, pp. 15–47.
45. G. Orsini, A. Ruggeri, Jr, A. Mazzoni, V. Papa, M. Piccirilli, M. Falconi, R. Di Lenarda and L. Breschi, "Immunohistochemical identification of type I and type III collagen and chondroitin sulphate in human pre-dentine: a correlative FEI-SEM/TEM study," *Int. Endod. J.*, vol. 40, pp. 669–678, Sep, 2007.
46. C. Guidry and F. Grinnell, "Studies on the mechanism of hydrated collagen gel reorganization by human skin fibroblasts," *J. Cell. Sci.*, vol. 79, pp. 67–81, Nov, 1985.
47. E. A. Sander and V. H. Barocas, "Biomimetic collagen tissues: collagenous tissue engineering and other applications," in *Collagen: Structure and Mechanics*, P. Fratzl, Ed. New York: Springer, 2008, pp. 475–504.
48. P. D. Yurchenco and J. C. Schittny, "Molecular architecture of basement membranes," *FASEB J.*, vol. 4, pp. 1577–1590, Apr 1, 1990.
49. P. D. Yurchenco, "Assembly of basement membranes," *Ann. NY Acad. Sci.*, vol. 580, pp. 195–213, 1990.
50. J. H. Miner and P. D. Yurchenco, "Laminin functions in tissue morphogenesis," *Annu. Rev. Cell Dev. Biol.*, vol. 20, pp. 255–284, 2004.
51. R. B. Colvin, "Fibronectin in wound healing," in *Fibronectin*, D. F. Mosher, Ed. Newyork: Academic press, pp. 213–254, 1989.

52. I. Wierzbicka-Patynowski and J. E. Schwarzbauer, "The ins and outs of fibronectin matrix assembly," *J. Cell. Sci.*, vol. 116, pp. 3269–3276, Aug 15, 2003.
53. M. K. Magnusson and D. F. Mosher, "Fibronectin: structure, assembly, and cardiovascular implications," *Arterioscler. Thromb. Vasc. Biol.*, vol. 18, pp. 1363–1370, Sep, 1998.
54. W. E. Fowler, R. R. Hantgan, J. Hermans and H. P. Erickson, "Structure of the fibrin protofibril," *Proc. Natl. Acad. Sci. USA*, vol. 78, pp. 4872–4876, Aug, 1981.
55. J. W. Weisel and C. Nagaswami, "Computer modeling of fibrin polymerization kinetics correlated with electron microscope and turbidity observations: clot structure and assembly are kinetically controlled," *Biophys. J.*, vol. 63, pp. 111–128, Jul, 1992.
56. M. W. Mosesson, "Fibrinogen and fibrin polymerization: appraisal of the binding events that accompany fibrin generation and fibrin clot assembly," *Blood Coagul. Fibrin.*, vol. 8, pp. 257–267, Jul, 1997.
57. M. R. Neidert, E. S. Lee, T. R. Oegema and R. T. Tranquillo, "Enhanced fibrin remodeling in vitro with TGF-beta1, insulin and plasmin for improved tissue-equivalents," *Biomaterials*, vol. 23, pp. 3717–3731, Sep, 2002.
58. E. D. Grassl, T. R. Oegema and R. T. Tranquillo, "Fibrin as an alternative biopolymer to type-I collagen for the fabrication of a media equivalent," *J. Biomed. Mater. Res.*, vol. 60, pp. 607–612, Jun 15, 2002.
59. V. H. Barocas, A. G. Moon and R. T. Tranquillo, "The fibroblast-populated collagen microsphere assay of cell traction force – Part 2: Measurement of the cell traction parameter," *J. Biomech. Eng.*, vol. 117, pp. 161–170, May, 1995.
60. M. L. Gardel, J. H. Shin, F. C. MacKintosh, L. Mahadevan, P. A. Matsudaira and D. A. Weitz, "Scaling of F-actin network rheology to probe single filament elasticity and dynamics," *Phys. Rev. Lett.*, vol. 93, pp. 188102, Oct 29, 2004.
61. O. Lieleg, M. M. Claessens, C. Heussinger, E. Frey and A. R. Bausch, "Mechanics of bundled semiflexible polymer networks," *Phys. Rev. Lett.*, vol. 99, pp. 088102, Aug 24, 2007.
62. P. L. Chandran and V. H. Barocas, "Microstructural mechanics of collagen gels in confined compression: poroelasticity, viscoelasticity, and collapse," *J. Biomech. Eng.*, vol. 126, pp. 152–166, Apr, 2004.
63. D. C. Morse, "Viscoelasticity of concentrated isotropic solutions of semiflexible polymers. 1. Model and stress tensor," *Macromolecules*, vol. 31, pp. 7030–7043, 1998.
64. D. C. Morse, "Viscoelasticity of concentrated isotropic solutions of semiflexible polymers. 2. Linear response," *Macromolecules*, vol. 31, pp. 7044–7067, 1998.
65. D. M. Knapp, V. H. Barocas, A. G. Moon, K. Yoo, L. R. Petzold and R. T. Tranquillo, "Rheology of reconstituted type I collagen gel in confined compression," *J. Rheol.*, vol. 41, pp. 971–993, 1997.
66. C. F. Schmidt, M. Baermann, G. Isenberg and E. Sackmann, "Chain dynamics, mesh size, and diffusive transport in networks of polymerized actin: a quasielastic light scattering and microfluorescence study," *Macromolecules*, vol. 22, pp. 3638–3649, 1989.
67. B. Wagner, R. Tharmann, I. Haase, M. Fischer and A. R. Bausch, "Cytoskeletal polymer networks: the molecular structure of cross-linkers determines macroscopic properties," *Proc. Natl. Acad. Sci. USA*, vol. 103, pp. 13974–13978, Sep 19, 2006.
68. J. H. Shin, M. L. Gardel, L. Mahadevan, P. Matsudaira and D. A. Weitz, "Relating microstructure to rheology of a bundled and cross-linked F-actin network in vitro," *Proc. Natl. Acad. Sci. USA*, vol. 101, pp. 9636–9641, Jun 29, 2004.
69. P. A. Janmey, M. E. McCormick, S. Rammensee, J. L. Leight, P. C. Georges and F. C. MacKintosh, "Negative normal stress in semiflexible biopolymer gels," *Nat. Mater.*, vol. 6, pp. 48–51, Jan, 2007.
70. C. Storm, J. J. Pastore, F. C. MacKintosh, T. C. Lubensky and P. A. Janmey, "Nonlinear elasticity in biological gels," *Nature*, vol. 435, pp. 191–194, May 12, 2005.

71. B. A. Roeder, K. Kokini, J. E. Sturgis, J. P. Robinson and S. L. Voytik-Harbin, "Tensile mechanical properties of three-dimensional type I collagen extracellular matrices with varied microstructure," *J. Biomech. Eng.*, vol. 124, pp. 214–222, Apr, 2002.
72. T. T. Tower, M. R. Neidert and R. T. Tranquillo, "Fiber alignment imaging during mechanical testing of soft tissues," *Ann. Biomed. Eng.*, vol. 30, pp. 1221–1233, Nov–Dec, 2002.
73. P. L. Chandran and V. H. Barocas, "Affine versus non-affine fibril kinematics in collagen networks: theoretical studies of network behavior," *J. Biomech. Eng.*, vol. 128, pp. 259–270, Apr, 2006.
74. P. L. Chandran and V. H. Barocas, "Deterministic material-based averaging theory model of collagen gel micromechanics," *J. Biomech. Eng.*, vol. 129, pp. 137–147, Apr, 2007.
75. B. A. Roeder, K. Kokini, J. P. Robinson and S. L. Voytik-Harbin, "Local, three-dimensional strain measurements within largely deformed extracellular matrix constructs," *J. Biomech. Eng.*, vol. 126, pp. 699–708, Dec, 2004.
76. Y. C. Fung, *Biomechanics: Mechanical Properties of Living Tissues*, 2nd ed. New York: Springer-Verlag, 1993, 568 pp.
77. L. Krishnan, J. A. Weiss, M. D. Wessman and J. B. Hoying, "Design and application of a test system for viscoelastic characterization of collagen gels," *Tissue Eng.*, vol. 10, pp. 241–252, Jan–Feb, 2004.
78. T. S. Girton, V. H. Barocas and R. T. Tranquillo, "Confined compression of a tissue-equivalent: collagen fibril and cell alignment in response to anisotropic strain," *J. Biomech. Eng.*, vol. 124, pp. 568–575, Oct, 2002.
79. V. C. Mow, S. C. Kuei, W. M. Lai and C. G. Armstrong, "Biphasic creep and stress relaxation of articular cartilage in compression: Theory and experiments," *J. Biomech. Eng.*, vol. 102, pp. 73–84, 1980.
80. J. K. Suh, Z. Li and S. L. Woo, "Dynamic behavior of a biphasic cartilage model under cyclic compressive loading," *J. Biomech.*, vol. 28, pp. 357–364, 1995.
81. G. C. Wood and M. K. Keech, "The formation of fibrils from collagen solutions. 1. The effect of experimental conditions: kinetic and electron-microscope studies," *Biochem. J.*, vol. 75, pp. 588–598, Jun, 1960.
82. C. B. Raub, V. Suresh, T. Krasieva, J. Lyubovitsky, J. D. Mih, A. J. Putnam, B. J. Tromberg and S. C. George, "Noninvasive assessment of collagen gel microstructure and mechanics using multiphoton microscopy," *Biophys. J.*, vol. 92, pp. 2212–2222, Mar 15, 2007.
83. A. M. Stein, D. A. Vader, L. M. Jawerth, D. A. Weitz and L. M. Sander, "An algorithm for extracting the network geometry of 3d collagen gels," *J. Microscopy*, vol. 232, pp. 463–475, 2008.
84. A. M. Stein, D. A. Vader, D. A. Weitz and L. M. Sander, "The micromechanics of collagen I gels," *Arxiv*, vol. 0807.2805, 2008.
85. S. Hsu, A. M. Jamieson, and J. Blackwell, "Viscoelastic studies of extracellular matrix interactions in a model native collagen gel system," Biorheology, vol. 31, pp. 21–36, 1994.
86. R. M. Kuntz and W. M. Saltzman, "Neutrophil motility in extracellular matrix gels: mesh size and adhesion affect speed of migration," Biophys. J., vol 72, pp. 1472–1480, 1997.
87. J. D. Humphrey, *Cardiovascular Solid Mechanics*. New York: Springer, 2002, 776pp.
88. C. F. Burgoyne, J. C. Downs, A. J. Bellezza and R. T. Hart, "Three-dimensional reconstruction of normal and early glaucoma monkey optic nerve head connective tissues," *Invest. Ophthalmol. Vis. Sci.*, vol. 45, pp. 4388–4399, Dec, 2004.
89. P. Bajcsy, S. C. Lee, A. Lin and R. Folberg, "Three-dimensional volume reconstruction of extracellular matrix proteins in uveal melanoma from fluorescent confocal laser scanning microscope images," *J. Microsc.*, vol. 221, pp. 30–45, Jan, 2006.
90. P. J. Elbischger, H. Bischof, P. Regitnig and G. A. Holzapfel, "Automatic analysis of collagen fiber orientation in the outermost layer of human arteries," *Pattern Anal. Appl.*, vol. 7, pp. 269–284, 2004.

91. B. van Rietbergen, S. Majumdar, W. Pistoia, D. C. Newitt, M. Kothari, A. Laib and P. Ruegsegger, "Assessment of cancellous bone mechanical properties from micro-FE models based on micro-CT, pQCT and MR images," *Technol. Health Care*, vol. 6, pp. 413–420, Dec, 1998.
92. B. Van Rietbergen, R. Muller, D. Ulrich, P. Ruegsegger and R. Huiskes, "Tissue stresses and strain in trabeculae of a canine proximal femur can be quantified from computer reconstructions," *J. Biomech.*, vol. 32, pp. 443–451, Apr, 1999.
93. H. H. Bayraktar and T. M. Keaveny, "Mechanisms of uniformity of yield strains for trabecular bone," *J. Biomech.*, vol. 37, pp. 1671–1678, Nov, 2004.
94. R. Muller, A. Nazarian, P. Schneider, M. Stauber, P. Thurner, G. H. van Lenthe and R. Voide, "Functional micro-imaging at the interface of bone mechanics and biology," in *Mechanics of Biological Tissue* G. A. Holzapfel and R. W. Odgen, Eds. Germany: Springer-Verlag, 2006, pp. 473–487.
95. C. Hitzenberger, E. Gotzinger, M. Sticker, M. Pircher and A. Fercher, "Measurement and imaging of birefringence and optic axis orientation by phase resolved polarization sensitive optical coherence tomography," *Opt. Exp.*, vol. 9, pp. 780–790, 2001.
96. M. Brezinski and J. Fujimoto, "Optical coherence tomography: high-resolution imaging innontransparent tissue," *IEEE J. Select. Topics Quant. Electron.*, vol. 5, pp. 1185–1192, 1999.
97. J. Rogowska, N. A. Patel, J. G. Fujimoto and M. E. Brezinski, "Optical coherence tomographic elastography technique for measuring deformation and strain of atherosclerotic tissues," *Heart*, vol. 90, pp. 556–562, May, 2004.
98. M. K. Al-Qaisi and T. Akkin, "Polarization-sensitive optical coherence tomography based on polarization-maintaining fibers and frequency multiplexing," *Opt. Express*, vol. 16, pp. 13032–13041, 2008.
99. W. Sun, A. Darling, B. Starly and J. Nam, "Computer-aided tissue engineering: overview, scope and challenges," *Biotechnol. Appl. Biochem.*, vol. 39, pp. 29–47, Feb, 2004.
100. P. Friedl and E. Brocker, "Biological confocal reflection microscopy: reconstruction of three-dimensional extracellular matrix, cell migration, and matrix reorganization," in *Image Analysis: Methods and Applications* D. Hader, Ed.. Boca Raton: CRC Press, 2001, pp. 9–22.
101. S. L. Voytik-Harbin, B. A. Roeder, J. E. Sturgis, K. Kokini and J. P. Robinson, "Simultaneous mechanical loading and confocal reflection microscopy for three-dimensional microbiomechanical analysis of biomaterials and tissue constructs," *Microsc. Microanal.*, vol. 9, pp. 74–85, Feb. 2003.
102. J. Liu, G. H. Koenderink, K. E. Kasza, F. C. Mackintosh and D. A. Weitz, "Visualizing the strain field in semiflexible polymer networks: strain fluctuations and nonlinear rheology of F-actin gels," *Phys. Rev. Lett.*, vol. 98, pp. 198304, May 11, 2007.
103. P. D. Yurchenco and H. Furthmayr, "Self-assembly of basement membrane collagen," *Biochemistry*, vol. 23, pp. 1839–1850, Apr 10, 1984.
104. A. S. Charonis, E. C. Tsilibary, P. D. Yurchenco and H. Furthmayr, "Binding of laminin to type IV collagen: a morphological study," *J. Cell Biol.*, vol. 100, pp. 1848–1853, Jun, 1985.
105. D. M. Shotton, B. E. Burke and D. Branton, "The molecular structure of human erythrocyte spectrin. Biophysical and electron microscopic studies," *J. Mol. Biol.*, vol. 131, pp. 303–329, Jun 25, 1979.
106. E. Katayama, "Quick-freeze deep-etch electron microscopy of the actin-heavy meromyosin complex during the in vitro motility assay," *J. Mol. Biol.*, vol. 278, pp. 349–367, May 1, 1998.
107. D. Overby, J. Ruberti, H. Gong, T. F. Freddo and M. Johnson, "Specific hydraulic conductivity of corneal stroma as seen by quick-freeze/deep-etch," *J. Biomech. Eng.*, vol. 123, pp. 154–161, Apr, 2001.
108. S. Erlandsen, C. Frethem and Y. Chen, "Field emission scanning electron microscopy (FESEM) Entering the 21st century: nanometer resolution and molecular topography of cell structure," *J. Histotechnol.*, vol. 23, pp. 249–259, 2000.

109. Y. Chen, L. Zardi and D. M. P. Peters, "high-resolution cryo-scanning electron microscopy study of the macromolecular structure of fibronectin fibrils," *Scanning*, vol. 19, pp. 349–355, 1997.
110. B. F. McEwen and M. Marko, "The emergence of electron tomography as an important tool for investigating cellular ultrastructure," *J. Histochem. Cytochem.*, vol. 49, pp. 553–564, May, 2001.
111. C. Baldock, C. J. Gilpin, A. J. Koster, U. Ziese, K. E. Kadler, C. M. Kielty and D. F. Holmes, "Three-dimensional reconstructions of extracellular matrix polymers using automated electron tomography," *J. Struct. Biol.*, vol. 138, pp. 130–136, Apr–May, 2002.
112. P. L. Kronick and M. S. Sacks, "Quantification of vertical-fiber defect in cattle hide by small-angle light scattering," *Connect. Tissue Res.*, vol. 27, pp. 1–13, 1991.
113. T. T. Tower and R. T. Tranquillo, "Alignment maps of tissues: I. Microscopic elliptical polarimetry," *Biophys. J.*, vol. 81, pp. 2954–2963, Nov, 2001.
114. K. L. Billiar and M. S. Sacks, "A method to quantify the fiber kinematics of planar tissues under biaxial stretch," *J. Biomech.*, vol. 30, pp. 753–756, Jul, 1997.
115. M. R. Neidert and R. T. Tranquillo, "Tissue-engineered valves with commissural alignment," *Tissue Eng.*, vol. 12, pp. 891–903, Apr, 2006.
116. B. C. Isenberg, C. Williams and R. T. Tranquillo, "Small-diameter artificial arteries engineered in vitro," *Circ. Res.*, vol. 98, pp. 25–35, Jan 6, 2006.
117. E. A. Sander, T. Stylianopoulos, R. T. Tranquillo and V. H. Barocas, "Image-based biomechanics of collagen-based tissue equivalents: multiscale models compared to fiber alignment predicted by polarimetric imaging," *IEEE Eng. Med. Biol. Mag.*, vol. 28, pp. 10–18, 2009.
118. C. Jhun, M.C. Evans, V. H. Barocas and R. T. Tranquillo, "Planar biaxial mechanical behavior of bioartificial tissues possessing prescribed fiber alignment," *J. Biomech. Eng.*, vol. 131:081006, 2009.
119. E. E. Underwood, *Quantitative Stereology*. Reading, MA: Addison-Wesley, 1970, 274pp.
120. C. Howard and M. Reed, *Unbiased Stereology: Three-Dimensional Measurement in Microscopy*. New York: Springer, 1998.
121. S. C. Cowin, "Wolff's law of trabecular architecture at remodeling equilibrium," *J. Biomech. Eng.*, vol. 108, pp. 83–88, Feb, 1986.
122. A. Odgaard, "Three-dimensional methods for quantification of cancellous bone architecture," *Bone*, vol. 20, pp. 315–328, 1997.
123. V. H. Barocas and R. T. Tranquillo, "An anisotropic biphasic theory of tissue-equivalent mechanics: the interplay among cell traction, fibrillar network deformation, fibril alignment, and cell contact guidance," *J. Biomech. Eng.*, vol. 119, pp. 137–145, May, 1997.
124. T. P. Harrigan and R. W. Mann, "Characterization of microstructural anisotropy in orthotropic materials using a second rank tensor," *J. Mat. Sci.*, vol. 19, pp. 761–767, 1984.
125. E. A. Sander and V. H. Barocas, "Comparison of 2D fiber network orientation measurement methods," *J. Biomed. Mater. Res. A.*, vol. 88, pp. 322–331, 2008.
126. W. J. Karlon, J. W. Covell, A. D. McCulloch, J. J. Hunter and J. H. Omens, "Automated measurement of myofiber disarray in transgenic mice with ventricular expression of ras," *Anat. Rec.*, vol. 252, pp. 612–625, Dec, 1998.
127. W. J. Karlon, P. P. Hsu, S. Li, S. Chien, A. D. McCulloch and J. H. Omens, "Measurement of orientation and distribution of cellular alignment and cytoskeletal organization," *Ann. Biomed. Eng.*, vol. 27, pp. 712–720, Nov–Dec, 1999.
128. M. Yoshigi, E. B. Clark and H. J. Yost, "Quantification of stretch-induced cytoskeletal remodeling in vascular endothelial cells by image processing," *Cytometry A.*, vol. 55, pp. 109–118, Oct, 2003.
129. W. M. Petroll, H. D. Cavanagh, P. Barry, P. Andrews and J. V. Jester, "Quantitative analysis of stress fiber orientation during corneal wound contraction," *J. Cell. Sci.*, vol. 104 (Pt 2), pp. 353–363, Feb, 1993.

130. B. Pourdeyhimi and R. Dent, "Measuring fiber orientation in nonwovens: Part II: Direct tracking," *Text. Res. J.*, vol. 66, pp. 747–753, 1996.
131. J. Wu, B. Rajwa, D. L. Filmer, C. M. Hoffmann, B. Yuan, C. Chiang, J. Sturgis and J. P. Robinson, "Automated quantification and reconstruction of collagen matrix from 3D confocal datasets," *J. Microsc.*, vol. 210, pp. 158–165, May, 2003.
132. J. Wu, B. Rajwa, D. L. Filmer, C. M. Hoffmann, B. Yuan, C. S. Chiang, J. Sturgis and J. P. Robinson, "Analysis of orientations of collagen fibers by novel fiber-tracking software," *Microsc. Microanal.*, vol. 9, pp. 574–580, Dec, 2003.
133. Y. Xia and K. Elder, "Quantification of the graphical details of collagen fibrils in transmission electron micrographs," *J. Microsc.*, vol. 204, pp. 3–16, Oct, 2001.
134. J. P. Marquez, "Fourier analysis and automated measurement of cell and fiber angular orientation distributions." *Int. J. Solids Struct.*, vol. 43, pp. 6413–6423, 2006.
135. E. Ghassemieh, M. Acar and H. Versteeg, "Microstructural analysis of non-woven fabric using scanning electron microscopy and image processing. Part 2: Application to hydroentangled fabrics." *Proc. Instn. Mech. Engrs. L.*, vol. 216, pp. 211–218, 2002.
136. B. Xu and L. Yu, "Determining fiber orientation distribution in nonwovens with hough transform techniques," *Text. Res. J.*, vol. 67, pp. 563–571, 1997.
137. A. Takahashi, R. Kita, T. Shinozaki, K. Kubota and M. Kaibara, "Real space observation of three-dimensional network structure of hydrated fibrin gel," *Coll. Polym. Sci.*, vol. 281, pp. 832–838, 2003.
138. R. C. Gonzalez, R. E. Woods and S. L. Eddins, *Digital Image Processing using MATLAB*. Upper Saddle River, NJ: Pearson Prentice Hall, 2004, 609pp.
139. L. J. Gibson and M. F. Ashby, *Cellular Solids: Structures and Properties*, 2nd ed. Oxford: Pergamon Press, 1997.
140. C. Heussinger and E. Frey, "Stiff polymers, foams, and fiber networks," *Phys. Rev. Lett.*, vol. 96, pp. 017802, Jan 13, 2006.
141. C. Heussinger and E. Frey, "Role of architecture in the elastic response of semiflexible polymer and fiber networks," *Phys. Rev. E. Stat. Nonlin Soft Matter Phys.*, vol. 75, pp. 011917, Jan, 2007.
142. B. Pourdeyhimi, R. Ramanathan and R. Dent, "Measuring fiber orientation in nonwovens: Part I: Simulation," *Text. Res. J.*, vol. 66, pp. 713–722, 1996.
143. D. A. Head, A. J. Levine and F. C. MacKintosh, "Distinct regimes of elastic response and deformation modes of cross-linked cytoskeletal and semiflexible polymer networks," *Phys. Rev. E. Stat. Nonlin Soft Matter Phys.*, vol. 68, pp. 061907, Dec, 2003.
144. M. Kellomaki, J. Astrom and J. Timonen, "Rigidity and dynamics of random spring networks," *Phys. Rev. Lett.*, vol. 77, pp. 2730–2733, Sep 23, 1996.
145. A. M. Sastry, C. Cheng and C. W. Wang, "Mechanics of stochastic fibrous networks," *J. Thermoplast. Compos. Mater.*, vol. 11, pp. 288–296, 1998.
146. C. W. Wang, L. Berhan and A. M. Sastry, "Structure, mechanics and failure of stochastic fibrous networks: Part I: Microscale considerations," *J. Eng. Mater. Technol.*, vol. 122, pp. 450–468, 2000.
147. P. R. Onck, T. Koeman, T. van Dillen and E. van der Giessen, "Alternative explanation of stiffening in cross-linked semiflexible networks," *Phys. Rev. Lett.*, vol. 95, pp. 178102, Oct 21, 2005.
148. B. Agoram and V. H. Barocas, "Coupled macroscopic and microscopic scale modeling of fibrillar tissues and tissue equivalents," *J. Biomech. Eng.*, vol. 123, pp. 362–369, Aug, 2001.
149. T. Stylianopoulos and V. H. Barocas, "Volume-averaging theory for the study of the mechanics of collagen networks," *Comput. Methods Appl. Mech. Eng.*, 2007.
150. T. Stylianopoulos and V. H. Barocas, "Multiscale, structure-based modeling for the elastic mechanical behavior of arterial walls," *J. Biomech. Eng.*, vol. 129, pp. 611–618, Aug, 2007.

151. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller, "Perspective on ''Equation of state calculations by fast computing machines'','' *J. Chem. Phys.*, vol. 21, pp. 1087–1092, 1953.
152. W. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.
153. M. Allen and D. Tildesley, *Computer Simulation of Liquids*. USA: Oxford University Press, 1989.
154. D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*. New York: Cambridge University Press, 2005.
155. F. J. Vesely, *Computational Physics: An Introduction*. New York: Kluwer Academic/Plenum Publishers, 2001.
156. J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton*. Sunderland, MA: Sinauer Associates, 2001, 367pp.
157. M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*. New York: Oxford University Press, 1986.
158. F. C. MacKintosh, J. Kas and P. A. Janmey, "Elasticity of semiflexible biopolymer networks," *Phys. Rev. Lett.*, vol. 75, pp. 4425–4428, Dec 11, 1995.
159. Y. Lanir, "Constitutive equations for fibrous connective tissues," *J. Biomech.*, vol. 16, pp. 1–12, 1983.
160. K. L. Billiar and M. S. Sacks, "Biaxial mechanical properties of the native and glutaraldehyde-treated aortic valve cusp: Part II – A structural constitutive model," *J. Biomech. Eng.*, vol. 122, pp. 327–335, Aug, 2000.
161. P. M. Pinsky, D. van der Heide and D. Chernyak, "Computational modeling of mechanical anisotropy in the cornea and sclera," *J. Cataract Refract. Surg.*, vol. 31, pp. 136–145, Jan, 2005.
162. Y. Lanir, "The rheological behavior of the skin: experimental results and a structural model," *Biorheology*, vol. 16, pp. 191–202, 1979.
163. T. M. Quinn and V. Morel, "Microstructural modeling of collagen network mechanics and interactions with the proteoglycan gel in articular cartilage," *Biomech. Model. Mechanobiol.*, vol. 6, pp. 73–82, Jan, 2007.
164. H. Hatami-Marbini and R. Picu, "Scaling of nonaffine deformation in random semiflexible fiber networks," *Phys. Rev. E*, vol. 77, pp. 62103, 2008.
165. A. M. Pizzo, K. Kokini, L. C. Vaughn, B. Z. Waisner and S. L. Voytik-Harbin, "Extracellular matrix (ECM) microstructural composition regulates local cell-ECM biomechanics and fundamental fibroblast behavior: a multidimensional perspective," *J. Appl. Physiol.*, vol. 98, pp. 1909–1921, May, 2005.
166. D. G. Hepworth, A. Steven-fountain, D. M. Bruce and J. F. Vincent, "Affine versus non-affine deformation in soft biological tissues, measured by the reorientation and stretching of collagen fibres through the thickness of compressed porcine skin," *J. Biomech.*, vol. 34, pp. 341–346, Mar, 2001.
167. C. Heussinger, B. Schaefer and E. Frey, "Nonaffine rubber elasticity for stiff polymer networks," *Phys. Rev. E. Stat. Nonlin Soft Matter Phys.*, vol. 76, pp. 031906, Sep, 2007.
168. Q. Wen, A. Basu, J. P. Winer, A. Yodh and P. A. Janmey, "Local and global deformations in a strain-stiffening fibrin gel," *New J. Phys.*, vol. 9, pp. 428, 2007.
169. D. Rodney, M. Fivel and R. Dendievel, "Discrete modeling of the mechanics of entangled materials," *Phys. Rev. Lett.*, vol. 95, pp. 108004, Sep 2, 2005.
170. J. C. Maxwell, "On the calculation of the equilibrium and stiffness of frames," *Philosophical Magazine*, vol. 27, pp. 598–604, 1864.
171. M. Wyart, H. Liang, A. Kabla and L. Mahadevan, "Elasticity of soft particles and colloids near Random Close Packing," *Arxiv*, vol. 0806.4653, 2008.

172. G. A. Buxton and N. Clarke, ""Bending to stretching" transition in disordered networks," *Phys. Rev. Lett.*, vol. 98, pp. 238103, Jun 8, 2007.
173. J. S. Palmer and M. C. Boyce, "Constitutive modeling of the stress–strain behavior of F-actin filament networks," *Acta Biomaterialia*, vol. 4, pp. 597–612, 2008.
174. J. E. Bischoff, E. M. Arruda and K. Grosh, "Finite element modeling of human skin using an isotropic, nonlinear elastic constitutive model," *J. Biomech.*, vol. 33, pp. 645–652, Jun, 2000.
175. I. Klapper and H. Qian, "Remarks on discrete and continuous large-scale models of DNA dynamics," *Biophys. J.*, vol. 74, pp. 2504–2514, 1998.
176. K. J. Bathe, "Finite Element Procedures," New Jersery: Prentice-Hall, 1996.
177. J. H. Shin, L. Mahadevan, P. T. So and P. Matsudaira, "Bending stiffness of a crystalline actin bundle," *J. Mol. Biol.*, vol. 337, pp. 255–261, Mar 19, 2004.
178. C. Heussinger, M. Bathe and E. Frey, "Statistical mechanics of semiflexible bundles of wormlike polymer chains," *Phys. Rev. Lett.*, vol. 99, pp. 048101, Jul 27, 2007.
179. M. Bathe, C. Heussinger, M. M. Claessens, A. R. Bausch and E. Frey, "Cytoskeletal bundle mechanics," *Biophys. J.*, vol. 94, pp. 2955–2964, Apr 15, 2008.
180. D. A. Head, F. C. MacKintosh and A. J. Levine, "Nonuniversality of elastic exponents in random bond-bending networks," *Phys. Rev. E. Stat. Nonlin Soft Matter Phys.*, vol. 68, pp. 025101, Aug, 2003.
181. J. Wilhelm and E. Frey, "Elasticity of stiff polymer networks," *Phys. Rev. Lett.*, vol. 91, pp. 108103, Sep 5, 2003.
182. E. M. Huisman, T. van Dillen, P. R. Onck and E. Van der Giessen, "Three-dimensional cross-linked F-actin networks: relation between network architecture and mechanical behavior," *Phys. Rev. Lett.*, vol. 99, pp. 208103, Nov 16, 2007.
183. C. Wang and A. Sastry, "Structure, mechanics and failure of stochastic fibrous networks: Part II—Network simulations and application," *J. Eng. Mater. Technol.*, vol. 122, pp. 460, 2000.
184. K. K. Brewer, H. Sakai, A. M. Alencar, A. Majumdar, S. P. Arold, K. R. Lutchen, E. P. Ingenito and B. Suki, "Lung and alveolar wall elastic and hysteretic behavior in rats: effects of in vivo elastase treatment," *J. Appl. Physiol.*, vol. 95, pp. 1926–1936, Nov, 2003.
185. R. Hill, "Elastic properties of reinforced solids: some theoretical principles," *J. Mech. Phys. Solids*, vol. 11, pp. 357–372, 1963.
186. L. J. Gibson, "Biomechanics of cellular solids," *J. Biomech.*, vol. 38, pp. 377–399, Mar, 2005.
187. L. J. Gibson, "The mechanical behaviour of cancellous bone," *J. Biomech.*, vol. 18, pp. 317–328, 1985.
188. S. Vajjhala, A. M. Kraynik and L. J. Gibson, "A cellular solid model for modulus reduction due to resorption of trabeculae in bone," *J. Biomech. Eng.*, vol. 122, pp. 511–515, 2000.
189. E. A. Sander, D. A. Shimko, K. C. Dee and E. A. Nauman, "Examination of apparent and cellular level properties of human vertebral cancellous bone using combined cellular solid models," *Biomech. Model. Mechanobiol.*, vol. 2, pp. 97–107, Nov, 2003.
190. E. A. Sander, J. C. Downs, R. T. Hart, C. F. Burgoyne and E. A. Nauman, "A cellular solid model of the lamina cribrosa: mechanical dependence on morphology," *J. Biomech. Eng.*, vol. 128, pp. 879–889, Dec, 2006.
191. M. J. Silva and L. J. Gibson, "The effects of non-periodic microstructure and defects on the compressive strength of two-dimensional cellular solids," *Int. J. Mech. Sci.*, vol. 39, pp. 549–563, 1997.
192. D. A. Puleo and A. Nanci, "Understanding and controlling the bone-implant interface," *Biomaterials*, vol. 20, pp. 2311–2321, 1999.
193. M. H. Schwartz, P. H. Leo and J. L. Lewis, "A microstructural model for the elastic response of articular cartilage," *J. Biomech.*, vol. 27, pp. 865–873, Jul, 1994.

194. R. G. Breuls, B. G. Sengers, C. W. Oomens, C. V. Bouten and F. P. Baaijens, "Predicting local cell deformations in engineered tissue constructs: a multilevel finite element approach," *J. Biomech. Eng.*, vol. 124, pp. 198–207, Apr, 2002.
195. S. Nemat-Nasser and M. Hori, *Micromechanics: Overall Properties of Heterogeneous Materials*, 2nd ed. Amsterdam: North Holland, 1999.
196. M. Oda and K. Iwashita, *Mechanics of Granular Materials: An Introduction*. Rotterdam: Aa Balkema, 1999.
197. T. Stylianopoulos, C. A. Bashur, A. S. Goldstein, S. A. Guelcher and V. H. Barocas, "Computational predictions of the tensile properties of electrospun fiber meshes: effect of fiber diameter and fiber orientation." *J. Mech. Behav. Biomed. Mat.*, vol. 1, pp. 326–333, 2008.
198. J. Soulhat, M. Buschmann and A. Shirazi-Adl, "A fibril-network-reinforced biphasic model of cartilage in unconfined compression," *J. Biomech. Eng.*, vol. 121, pp. 340, 1999.
199. D. A. Drew, "Averaged field equations for two-phase media," *Stud. Appl. Math.*, vol. L, pp. 133–166, 1971.
200. B. Hinner, M. Tempel, E. Sackmann, K. Kroy and E. Frey, "Entanglement, elasticity, and viscous relaxation of actin solutions," *Phys. Rev. Lett.*, vol. 81, pp. 2614–2617, 1998.
201. M. Claessens, R. Tharmann, K. Kroy and A. Bausch, "Microstructure and viscoelasticity of confined semiflexible polymer networks," *Nat. Phys.*, vol. 2, pp. 186, 2006.
202. O. Lieleg and A. Bausch, "Cross-linker unbinding and self-similarity in bundled cytoskeletal networks," *Phys. Rev. Lett.*, vol. 99, pp. 158105, 2007.
203. R. Tharmann, M. M. Claessens and A. R. Bausch, "Viscoelasticity of isotropically cross-linked actin networks," *Phys. Rev. Lett.*, vol. 98, pp. 088103, Feb 23, 2007.
204. M. L. Gardel, J. H. Shin, F. C. MacKintosh, L. Mahadevan, P. Matsudaira and D. A. Weitz, "Elastic behavior of cross-linked and bundled actin networks," *Science*, vol. 304, pp. 1301–1305, May 28, 2004.
205. P. A. Janmey, U. Euteneuer, P. Traub and M. Schliwa, "Viscoelastic properties of vimentin compared with other filamentous biopolymer networks," *J. Cell Biol.*, vol. 113, pp. 155–160, 1991.
206. J. Leterrier, J. Kas, J. Hartwig, R. Vegners and P. Janmey, "Mechanical effects of neurofilament cross-bridges. Modulation by phosphorylation, lipids, and interactions with F-actin," *J. Biol. Chem.*, vol. 271, pp. 15687, 1996.
207. N. Wang and D. Stamenovic, "Contribution of intermediate filaments to cell stiffness, stiffening, and growth," *Am. J. Physiol. – Cell Physiol.*, vol. 279, pp. 188–194, 2000.
208. N. Wang, J. P. Butler and D. E. Ingber, "Mechanotransduction across the cell surface and through the cytoskeleton," *Science*, vol. 260, pp. 1124–1127, May 21, 1993.
209. D. E. Ingber, "Opposing views on tensegrity as a structural framework for understanding cell mechanics," *J. Appl. Physiol.*, vol. 89, pp. 1663–1670, Oct, 2000.
210. M. F. Coughlin and D. Stamenovic, "A tensegrity model of the cytoskeleton in spread and round cells," *J. Biomech. Eng.*, vol. 120, pp. 770–777, 1998.
211. D. Stamenovic and D. Ingber, "Models of cytoskeletal mechanics of adherent cells," *Biomech. Model. Mechanobiol.*, vol. 1, pp. 95–108, 2002.
212. C. P. Brangwynne, F. C. MacKintosh, S. Kumar, N. A. Geisse, J. Talbot, L. Mahadevan, K. K. Parker, D. E. Ingber and D. A. Weitz, "Microtubules can bear enhanced compressive loads in living cells because of lateral reinforcement," *J. Cell Biol.*, vol. 173, pp. 733, 2006.
213. D. E. Discher, D. H. Boal and S. Boey, "Phase transitions and anisotropic responses of planar triangular nets under large deformation," *Phys. Rev. E*, vol. 55, pp. 4762–4772, 1997.
214. S. K. Boey, D. H. Boal and D. E. Discher, "Simulations of the erythrocyte cytoskeleton at large deformation. I. Microscopic models," *Biophys. J.*, vol. 75, pp. 1573–1583, 1998.

215. R. Waugh and E. A. Evans, "Thermoelasticity of red blood cell membrane," *Biophys. J.*, vol. 26, pp. 115–131, Apr, 1979.
216. D. E. Discher, D. H. Boal and S. K. Boey, "Simulations of the erythrocyte cytoskeleton at large deformation. II. Micropipette aspiration," *Biophys. J.*, vol. 75, pp. 1584–1597, 1998.
217. J. C. M. Lee, D. T. Wong and D. E. Discher, "Direct measures of large, anisotropic strains in deformation of the erythrocyte cytoskeleton," *Biophys. J.*, vol. 77, pp. 853–864, 1999.
218. J. Li, M. Dao, C. Lim and S. Suresh, "Spectrin-level modeling of the cytoskeleton and optical tweezers stretching of the erythrocyte," *Biophys. J.*, vol. 88, pp. 3707–3719, 2005.
219. X. Luo, T. Stylianopoulos, V. H. Barocas and M. S. Shephard, "Multiscale computation for bioartificial soft tissues with complex geometries," *Eng. Comput.*, vol. 25, pp. 87–95, 2008.
220. N. Stergiopulos, S. Vulliemoz, A. Rachev, J. J. Meister and S. E. Greenwald, "Assessing the homogeneity of the elastic properties and composition of the pig aortic media," *J. Vasc. Res.*, vol. 38, pp. 237–246, May–Jun, 2001.
221. M. Frisen, M. Magi, I. Sonnerup and A. Viidik, "Rheological analysis of soft collagenous tissue. Part I: theoretical considerations," *J. Biomech.*, vol. 2, pp. 13–20, Mar, 1969.
222. P. S. Robinson, "Development of a functional tissue-engineered heart valve replacement," PhD dissertation, University of Minnesota, 2007.
223. S. Thomopoulos, G. M. Fomovsky, P. L. Chandran and J. W. Holmes, "Collagen fiber alignment does not explain mechanical anisotropy in fibroblast populated collagen gels," *J. Biomech. Eng.*, vol. 129, pp. 642, 2007.
224. H. J. Burd, "A structural constitutive model for the human lens capsule," *Biomech. Model. Mechanobiol*, Jul 13, 2008.
225. F. S. A. Cavalcante, S. Ito, K. Brewer, H. Sakai, A. M. Alencar, M. P. Almeida, J. S. Andrade, A. Majumdar, E. P. Ingenito and B. Suki, "Mechanical interactions between collagen and proteoglycans: implications for the stability of lung tissue," *J. Appl. Physiol.*, vol. 98, pp. 672–679, 2005.
226. P. A. DiMilla, K. Barbee and D. A. Lauffenburger, "Mathematical model for the effects of adhesion and mechanics on cell migration speed." *Biophys. J.*, vol. 60, pp. 15, 1991.
227. N. Wang and D. E. Ingber, "Control of cytoskeletal mechanics by extracellular matrix, cell shape, and mechanical tension," *Biophys. J.*, vol. 66, pp. 2181–2189, 1994.
228. I. B. Bischofs, F. Klein, D. Lehnert, M. Bastmeyer and U. S. Schwarz, "Filamentous network mechanics and active contractility determine cell and tissue shape," *Biophys. J.*, vol. 95, pp. 3488–3496, 2008.
229. M. J. Buehler, "Hierarchical nanomechanics of collagen fibrils: Atomistic and molecular modeling," in *Collagen: Structure and Mechanics*, Anonymous. New York: Springer, 2008, pp. 175–248.
230. S. J. Eppell, B. N. Smith, H. Kahn and R. Ballarini, "Nano measurements with micro-devices: mechanical properties of hydrated collagen fibrils," *J. R. Soc. Interface*, vol. 3, pp. 117–121, Feb 22, 2006.
231. A. Desai and T. J. Mitchison, "Microtubule polymerization dynamics," *Annu. Rev. Cell Dev. Biol.*, vol. 13, pp. 83–117, 1997.
232. V. Bennett and D. Gilligan, "The spectrin-based membrane skeleton and micron-scale organization of the plasma membrane," *Annu. Rev. Cell Biol.*, vol. 9, pp. 27–66, 1993.
233. J. P. Collet, D. Park, C. Lesty, J. Soria, C. Soria, G. Montalescot, and J. W. Weisel, "Influence of fibrin network conformation and fibrin fiber diameter on fibrinolysis speed: dynamic and structural approaches by confocal microscopy," *Ateriosclerosis. Thromb. Vasc. Biol.*, vol. 20, pp. 1354–1361, 2000.

INDEX

T. Dumitrica (ed.), Trends in Computational Nanomechanics, 603–620.
DOI 10.1007/978-1-4020-9785-0,