From atomistic simulation towards multiscale modelling of materials

2002 J. Phys.: Condens. Matter 14 2859

(http://iopscience.iop.org/0953-8984/14/11/306)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **14** (2002) 2859–2876 PII: S0953-8984(02)31183-4

From atomistic simulation towards multiscale modelling of materials

Risto M Nieminen

COMP/Laboratory of Physics, Helsinki University of Technology, PO Box 1100, 02015 HUT, Finland

E-mail: rni@fyslab.hut.fi

Received 23 November 2001, in final form 14 January 2002
Published 8 March 2002
Online at stacks.iop.org/JPhysCM/14/2859

Abstract

This paper discusses some current trends in computational materials science, especially the striving to forge links between modelling activities at various length and timescales. At the atomistic scale, methods based on quantum mechanical, especially density-functional, theories for electronic properties link to atomic/molecular dynamics and kinetic Monte Carlo simulations. Coarse graining leads to lattice-gas and cellular automata, and eventually to continuum equations solved by finite-element and finite-difference techniques. As examples of hierarchical modelling of materials, the paper describes recent work on anisotropic chemical etching of silicon, irradiation processing of fullerenes, oxygen clustering in silicon and self-diffusion in the compound semiconductor GaSb.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Sustainable economies with responsible utilization of natural resources and energy have to be based on advanced technologies, which, in turn, depend on advances in materials performance and on developing new materials with new, superior properties.

Modern materials science is a truly interdisciplinary endeavour, based on three fundamental activities. The *manufacture* and *processing* of materials uses the ingredients provided by the periodic table and naturally existing chemical precursors, together with a multitude of different physical, chemical and biological techniques, to make different materials in seemingly infinite varieties. The physical and chemical *characterization* of materials is made possible by a huge array of sophisticated techniques, with often atomic-scale resolution and extreme sensitivity. The *modelling* and *simulation* of materials are based on the development of theories and computational methods, and range from the quantum physics of atomic-scale phenomena to the continuum descriptions of macroscopic behaviour.

The realm of materials is vast, ranging from the inorganic world of metallic, semiconducting and insulating materials to organic polymers as well as biologically relevant and biomimetic materials and structures. Their physical, chemical and biological properties vary enormously, from ultrahard solids to soft tissues and DNA strands, from ceramic superconductors to organic semiconductors, from liquid ferromagnets to amorphous insulators. The structural variety is unbounded, and can be accessed with sophisticated manipulation and processing techniques.

The infinite number of possibilities in materials research underscores the importance of predictive theory and modelling. The community of materials theorists has traditionally been divided into subcommunities largely defined by the length and timescales of interest to them. In the *macroscopic* spatial regime from millimetres to metres, modelling has typically developed around continuum equations solved by finite-element (FE) and finite-difference techniques. In the *mesoscopic* regime from micrometres to millimetres, phenomenological approaches have developed around stochastic methods for the material's microstructure such as grain boundaries and dislocations. In the *microscopic*, Ångström-scale regime physicists and quantum chemists have based their work on the Schrödinger equation and other expressions of quantum mechanics of interacting electrons and atomic nuclei. The rapid increase in computational capabilities, in terms of both raw computing power and new algorithms, has enabled spectacular developments in each of these regimes. These entail increased complexity in its various forms: more degrees of freedom, more complicated boundary conditions and geometries, more nonlinearities, longer timescales, further away from equilibrium.

The challenge now faced by theorists is to bridge the different length (and time) scales to a more general framework, which has been coined as *multiscale modelling*. One should be able to move, as seamlessly as possible, from one scale to another so that the calculated parameters, properties and other numerical information calculated can be efficiently transferred across scales. Although all the information is in principle available at the finest (microscopic) level on description, it would be mindless and totally impractical to use it for macroscopic phenomena. At each length scale there are emergent phenomena which are best described by new, coarse-grained equations, which eliminate the unnecessary detail and emphasize the emergent properties. The challenge is how to optimally convey the information through appropriate 'hand-shaking' of the different regimes.

In very general terms, one can imagine two different answers to this challenge. In the *concurrent* approach one constructs a universal description (e.g. Hamiltonian) valid at all relevant scales, but so that the redundant degrees of freedom can be efficiently 'integrated out' for the actual treatment at one scale. For example, the Newton equations of motion for atomic-scale entities should be (reversibly) convertible to kinetic equations or lattice-gas automata (see below) for density and velocity fields. This approach resembles the multigrid techniques used for partial differential equations or the general renormalization-group idea of statistical physics. While intellectually most appealing, this approach has not yet been developed into a general scheme and will not be further discussed here. The reader is referred to interesting and important recent articles [1, 2].

In the *hierarchical* approach one carries out the computational simulation at one scale and extracts such quantities that can be used to define the parameters of the model operative on the adjoining (usually longer) scale. For example, the energy barriers, vibrational frequencies and entropies for atomic motion can be calculated at the microscopic level and then transferred to a kinetic simulation as discrete jump probabilities or as coefficients to reaction-diffusion-type equations.

The properties of materials and structures made thereof should ultimately be explainable in terms of their constituent atoms and their mutual interactions, and their motion at finite

temperatures and under external forces. Below, the basic steps from the microscopic towards the macroscopic are briefly described, and a few examples from recent research then follow.

2. Hierarchical modelling of materials phenomena

2.1. Atomistic calculations: density-functional theory

Density-functional theory (DFT) [3] is the undisputed workhorse for quantum mechanical atomistic calculations. DFT transforms the complex many-body problem of interacting electrons and nuclei into a coupled set of one-particle (Kohn–Sham) equations, which are computationally much more manageable. The theory allows parameter-free calculations of all ground-state physical observables, such as charge and spin densities, total energies and many related quantities, such as bonding distances, elastic moduli, vibrational frequencies, defect and surface energies, migration barriers, reaction energies, magnetic moments etc. There are numerous popular implementations of DFT to large-scale calculations of materials properties, for example those using plane-wave basis functions and pseudopotentials for valence electrons or those using linearized methods for all-electron calculations. System sizes of up to several hundred atoms are feasible. While the simple local-(spin-) density approximation (L(S)DA) is robust and provides often surprising quantitative accuracy, there are also workable methods beyond L(S)DA which recover aspects of the nonlocal nature of electronic exchange and correlation interactions. These methods can lead to a numerical accuracy similar that of full quantum calculations (configuration interactions, coupled cluster, quantum Monte Carlo (MC)) for small systems. The latter methods are currently computationally too heavy for large system sizes. However, one should note the important recent development towards a linear-scaling ('order-N') formulation of the quantum MC method [4]. The static formulation of DFT enables also 'first-principles' molecular dynamics simulations, where the interatomic forces are calculated for the adiabatic motion of nuclei from the electronic degrees of freedom, utilizing the Hellmann–Feynman theorem.

The density-functional approach can be generalized to the *time-dependent* case [5] as well. In the linear-response regime, such properties as photoabsorption spectra or frequency-dependent electromagnetic susceptibilities can be attacked. While the construction of the exchange–correlation functional for time-dependent calculations is still a challenge, the simple adiabatic local-density approximation (TDLDA), which implies an instantaneous response dependent only on the local electron density, has turned out to be surprisingly accurate for many purposes.

The time-dependent formulation also allows an approach to address the question of *excitation* energies, for which the ground-state Kohn–Sham eigenvalues often give a poor estimate. For finite systems such an approach has been shown to be quite accurate [5], comparable to more laborious quasiparticle 'self-energy' methods. In the case of strong external perturbations, such as laser dissociation of molecules, DFT provides a working scheme to simulate the full nonlinear response of the combined system of electrons and nuclei.

The numerical solution of the Kohn–Sham equations using real-space (RS) methods [6] instead of basis sets is currently an active research topic. Real-space grids can offer several benefits. Firstly, RS methods can in principle be used in both pseudopotential and all-electron calculations. Secondly, and more importantly in the present context, systems containing different length scales (e.g. nanostructures, surfaces etc) can be treated economically as one need not waste many grid points in empty regions. Also different types of boundary condition (free or periodic) are easily implemented. Finally, RS methods can can be efficiently adapted to parallel computing through domain decomposition. Multigrid techniques enable substantial

speedups in the convergence of RS methods, which are now becoming competitive with more conventional basis-set approaches such as Fourier series (plane-wave methods).

2.2. Moving atoms: molecular dynamics simulations

Within the Born–Oppenheimer (BO) approximation, the total ground-state energy associated with the electronic degrees of freedom defines the potential energy hypersurface for the atomic motion, which is classical and controlled by the nuclear masses. Exceptions are the lightest elements, hydrogen and helium, for which quantum mechanical tunnelling plays a role. The interatomic forces can be obtained as derivatives of the potential energy hypersurface with respect to nuclear coordinates. The low-lying excitations of the nuclear subsystem are quantized phonons, which can be accurately addressed by density-functional perturbation theory [7].

At high temperatures, the atomic motion becomes entirely classical. The atomic excursions from their equilibrium positions increase with increasing temperature. The motion becomes anharmonic and eventually leads to diffusive motion characterized by hops over barriers separating equivalent positions in the potential energy landscape. The atomic motion can be modelled by solving Newton's equations of motion. This approach goes under the generic name of *molecular dynamics simulations* [8], and has become a popular and powerful way to investigate the complicated, collective processes associated with atomic motion. Various thermodynamical ensembles (constant total energy—microcanonical, constant temperature—canonical, constant volume, constant pressure) can be simulated using, if necessary, auxiliary variables for atomic velocity scaling and unit-cell dimensions.

As molecular dynamics simulation works in 'real time', i.e. the computational timestep (typically less than 10^{-15} s) is dictated by the physical constants, following a particular physical event may be painfully slow and require a huge number of computational timesteps. This is the problem of *rare events*, such as a thermally activated jump over a migration barrier at low temperatures, which makes the brute force application of molecular dynamics simulation unpractical for such cases. Ingenious schemes have been proposed [9] to overcome the rare-event bottleneck while preserving the deterministic nature of molecular dynamics simulation.

The most direct, parameter-free approach to molecular dynamics simulation is to calculate the forces from first principles, i.e. by evaluating the electronic total energy and interatomic forces at each timestep. This can be done either 'on the fly' through the Car–Parrinello algorithm [10] (which updates the electronic and ionic degrees of freedom in unison in the vicinity of the BO surface) or through direct minimization of the electronic total energy on the BO surface (adiabatic molecular dynamics).

First-principles molecular dynamics simulations are computationally demanding and still limited to rather modest system sizes and short time sequences. The accurate evaluation of the forces also puts stringent requirements on the computational techniques used. Most of the practical implementations of first-principles molecular dynamics methods use Fourier techniques (plane-wave basis sets) for the electronic degrees of freedom in conjunction with pseudopotentials to fold out the inert core-electron states. Plane-wave methods have the advantage that the spatial resolution is uniform, i.e. independent of the nuclear positions, which enables accurate force evaluations. Even with the fastest (parallel) algorithms and computers, the calculations are typically limited to a few hundred electrons and atoms in size and to a few nanoseconds in length.

The molecular dynamics simulations can be speeded up at the expense of 'first-principles' accuracy of the force evaluations. Several schemes utilizing minimal basis sets in the form of local orbitals in the spirit of the tight-binding approximation have been introduced [11], with varying degrees of accuracy and portability. The force evaluations in such schemes are

usually faster than the first-principles methods by at least one order of magnitude. Schemes utilizing parametrized classical interactions in analytic form are naturally fastest. Such schemes range from simple pairwise interatomic potentials of insulating solids to angle-dependent models for covalent solids and to sophisticated many-atom force fields designed for organic molecules [12].

2.3. Moving atoms: kinetic Monte Carlo methods

Another possible approach to simulating atomic motion is provided by stochastic simulation by the MC method [13]. The basic idea is very simple. The potential energy hypersurface is first evaluated for all relevant atomic configurations. These include not only (meta)stable configurations but also those associated with saddle points. A typical saddle point configuration is one where an atom is at the transition state, at the barrier separating two nearby valleys in the hypersurface. A complete mapping of the hypersurface is of course impossible. Thus physical intuition is needed to divide the atoms into those actively moving during various processes of interest and those that are mere 'spectators' during the events. Identification of the barriers is a nontrivial task, and algorithms such as the 'nudged elastic band' (NEB) [14], the 'locally activated Monte Carlo' (LAMC) [15] or the 'tossing ropes' [16] are needed. The relative probabilities of different atomic movements are proportional to Boltzmann factors involving the energy differences between the transition state and the initial state.

In the kinetic Monte Carlo (KMC) simulation, one replaces the short-time dynamics of the system by discrete hops in a network. This is done by mapping the potential energy surface to a grid, where the grid points are associated with minima in the potential energy landscape. The possible initial and final states for atomic configurations are classified, typically in terms of 'atomic neighbourhoods', which largely determine the energy of a given configuration. The number, position and chemical identity of neighbouring atoms are used to define the class of the configuration.

Given the classified configurations, one defines the transition probability W_{fi} from an initial state i to a final state f for all possible transitions between the two. These can include not only single-atom events but also concerted events, i.e. simultaneous movements of several atoms. For a event class labelled by k, the transition rate is

$$\Gamma^k = W(f,i) = \Gamma_0^{fi} \exp\left(-\frac{E_{TS}^{fi} - E_i}{k_B T}\right) \tag{1}$$

where Γ_0^{fi} is a prefactor, E_{TS}^{fi} the total energy in the transition state for process k and E_i the energy of the initial state. The prefactor depends on entropic factors associated with the possible vibrational modes, and can be explicitly evaluated within the classical and harmonic approximation. Thermal equilibrium can be ensured by invoking the detailed-balance condition.

The KMC simulation is efficiently executed with the Bortz–Kalos–Lebowitz [17] algorithm. One first evaluates the total rate R by summing Γ^k over all possible processes. Secondly, three random numbers ρ_1, ρ_2 and ρ_3 are drawn from the interval [0,1]. One of them, say ρ_1, is used to determine the class k for the event chosen:

$$\sum_{j=0}^{k} \Gamma^j \leqslant \rho_1 R \leqslant \sum_{j=0}^{k-1} \Gamma^j. \tag{2}$$

The second random number ρ_2 is used to choose a particular event from the class k, and the third random number ρ_3 is used to advance the 'clock' through

$$t = t - \ln(\rho_3)/R. \tag{3}$$

The procedure can obviously be generalized for nonequilibrium dynamics controlled by arbitrary rate constants. KMC simulations are closely related to lattice-gas and cellular automaton models [18], where the updating of the local variables (configurations) follows rules written in terms of the neighbourhoods of each atom. The updating can take place either sequentially or in parallel. In the continuous cellular automaton scheme, the condition of discrete occupation numbers (e.g. that an atom either occupies a lattice site (occupation $= 1$) or the site is empty (occupation $= 0$)) is released. A continuous variable ranging from zero to unity is used to characterize the 'mass' or 'occupation state' of a lattice site. Starting from a site initially occupied by an atom, this continuous variable is reduced by an amount corresponding to its net removal probability. When the continuous variable reaches the value zero, the atom is definitely removed. This mode of KMC methods is particularly useful in simulations of surface etching, an example of which is discussed below.

2.4. From molecular dynamics to continuum equations

The coupling of the atomic-scale description of molecular dynamics to mesoscale and eventually macroscale models is absolutely essential to successful realization of true multiscale modelling. Let us consider the case of mechanical properties of solids as an example, including fracture and crack propagation under external load. Even though molecular dynamics simulations can currently run with hundreds of millions of atoms (with classical potentials), these calculations cannot properly represent environment of a dynamical system in the meso- and macroscales. The important issue is the flow of mechanical energy into and out of the system, rather than the statistical physics of the atomistic assembly (which can be modelled by a relatively small sample). Much of the important physics is in long-range interactions, such as the elastic field. Thus it is natural to envisage embedding the molecular-dynamics region in a continuum mechanical description. Far away from the central region (e.g. the tip of the propagating crack) the atoms are displaced only slightly from their equilibrium positions, and (linear) elasticity theory is expected to work well. The *FE method* [19] is the method of choice for this region. The FE algorithms for continuum equations are of course much more efficient computationally than molecular dynamics: only a small number of degrees of freedom are necessary. In crack propagation, much of the action is focused near the crack faces and the emitted dislocations. Far away from the crack, little is happening as atoms mainly vibrate around their equilibrium positions: a mean-field-type description is totally adequate. However, one cannot dispense with the distant region entirely. Molecular dynamics simulations show that long-wavelength pressure waves emanate from the crack tip. These waves must be properly propagated into the surrounding medium, which can only be accomplished by an appropriate embedding continuum.

The FE equations of motion can be derived straightforwardly from the Hamiltonian for linear elasticity theory, with the physical parameters contained in the elastic constant tensor $\boldsymbol{C}$ and the mass matrix $\boldsymbol{M}$. The crucial issue then is the 'handshaking' between the molecular dynamics region and the embedding FE region. This remains an active area of study, and several schemes have been proposed [20].

2.5. From kinetic Monte Carlo methods to rate equations

KMC schemes discussed above carry the spatial information defined in terms of discrete configurations of atoms through which the system evolves as a generalized random-walk process. This information can easily become overwhelming and difficult to handle and the number of possible configurations can become unmanageable. Instead of solving the full

stochastic differential equation (the 'master equation'), as the KMC methods do, one can define density fields for a set of relevant configurations, for example spatial densities $n_k(\boldsymbol{r})$ for atom clusters of size k. These density fields can then be linked with macroscopic mass currents, as for example in the case of simple chemical diffusion driven by density gradients. The diffusion current is

$$J_1 = -(Dn_1(\vec{r})/k_BT)\Delta\mu \tag{4}$$

where D is the phenomenological diffusion constant and μ the chemical potential. In the dilute limit

$$\mu = k_BT\ln(n_1(\vec{r})) \tag{5}$$

which leads to the Schmoluchowski equation and Fick's law. Combined with the constraint of mass conservation, the familiar *diffusion equation* is obtained, and a link between the microscopic random walk and the macroscopic diffusion model is established. Similarly, simple *rate-reaction equations* can be derived for spatially averaged quantities such as average concentrations of clusters with a given size. A famous example of such rate equations is Waite's formulation [21] for diffusion-limited reactions.

3. Anisotropic wet chemical etching of Si

Anisotropic wet chemical etching of silicon and other semiconductor materials is a key process in microelectronics processing. A typical example is the etching using a KOH solution, which leads to the removal of silicon atoms from the exposed surface. For smooth single-crystal surfaces, the etch rate is strongly dependent on the crystalline orientation. As an important application, the emerging microelectromechanical systems (MEMS) require the controlled extension of etching techniques from the essentially two-dimensional world of integrated microelectronic circuit design to three-dimensional structures of various actuators, sensors and oscillators. Predictive modelling of three-dimensional etching is thus an important challenge.

We have recently presented [22] a combination of *ab initio* level (i.e. free of adjustable parameters) atomic-scale calculations and MC/cellular automaton simulations, which provide a link between the macroscopic features of etched structures and the microscopic chemical reactions. In the atomistic model, the removal of a surface atom is controlled by its relative probability of forming a stable compound with the etchant radical approaching it from the solution. This probability depends on initial bonding of the surface atom to its neighbourhood (number of backbonds, nature of highest occupied and lowest unoccupied orbitals etc). Density-functional calculations can yield detailed information on the variety of possible initial configurations, reflecting the microscopic anisotropies of the underlying crystalline lattice. In particular, the relative energies of different configurations can be used to determine their relative probabilities.

In the discrete MC model, a surface atom is removed if a random number drawn from the interval [0, 1] is smaller than the relative removal probability. A single timestep of the evolution of the surface with N atoms is obtained after having tried N local updates (some of which have succeeded). After a successful local trial, the surface atom is removed and the state of the neighbourhood is updated. In the continuous MC scheme, a continuous variable ranging from unity to zero is used to characterize the occupation of each atomic site in the system. When a surface atom is chosen (at random), this continuous variable is reduced by an amount corresponding to the relative removal probability. When the continuous occupation variable eventually reaches the minimum value of zero, the atom is definitely removed. As

microscopic fluctuations of the system are suppressed, the continuous MC scheme starts to reflect the asymptotic macroscopic morphology of the evolving system with smaller system sizes than its discrete counterpart [22].

When considering the removal probabilities of surface atoms, it is necessary to consider local neighbourhoods which contain both nearest and next-nearest atoms. Let us assume that the dangling bonds of all surface atoms are terminated by either hydrogen or OH groups from the etching solution. DFT calculations show that this leads to weakening of the backbonds, arising from the electronegative pull exerted by the oxygen atom of the OH radical on the bonding electrons. As a result of this weakening, the probability of breaking a given bond depends on the number of OH groups attached to each of the atoms sharing that bond. For the diamond structure of silicon, there are 32 possible different configurations for a surface atom (19 configurations for an atom with three first neighbours, nine for an atom with two first neighbours and three for an atom with one first neighbour). These 32 cases correspond to eight different bond types. DFT cluster calculations are then carried out to obtain the removal energies for the different cases.

While the bond-breaking model outline above contains the effect of second neighbours, a detailed analysis shows that one also needs to distinguish between *direct* and *indirect* next-nearest neighbours. A direct second neighbour is linked to the target atom directly through a first neighbour, while an indirect second neighbour is linked through an indirect, longer path. Calculations for clusters with different numbers of indirect second neighbours show important differences. If the same bonds are broken upon removing a surface atom, the required energy is different when indirect second neighbours are present. It costs more energy to add an OH group to a silicon site with indirect neighbours. This leads to a reduced removal probability when compared with the case with no indirect neighbours.

The probability of removal of a surface atom is given by the modified Arrhenius expression

$$p = \exp\left(-\frac{\Delta E}{k_B T}\right) \tag{6}$$

where ΔE is defined as the excess of average bond energy $\bar{E}$ over a critical energy E_c:

$$\Delta E = k_B T \ln(1 + \mathrm{e}^{(\bar{E}-E_c)/k_B T}). \tag{7}$$

The critical energy E_c acts as a threshold below which bond breaking and removal occurs with essentially unit probability. The average energy per bond $\bar{E}$ is obtained from DFT calculations for the possible geometries with varying numbers of first and second (direct and indirect) neighbours.

Figure 1 shows the simulated 'wagon-wheel' pattern frequently used in studies of anisotropic etching, compared with the experimental mesoscale pattern (a). The system under investigation is the Si(100) wafer exposed to 10 wt% KOH solution at 75 °C. The comparison clearly shows the importance of indirect surface–atom interactions (c), and also demonstrates in a convincing way the power of modelling to draw a link between microscopic surface reactions and emerging, meso- and macroscale morphology.

4. Irradiation processing of fullerenes and carbon onions

The properties of carbon allotropes with various closed cage structures have been an active subject of research since the experimental discovery of fullerenes and nanotubes. The structure and energetics of fullerene spheres, tubes and nested forms known as carbon onions have been studied theoretically with a multitude of methods, including density-functional, Hartree–Fock and semi-empirical calculations of their electronic properties. Processing experiments

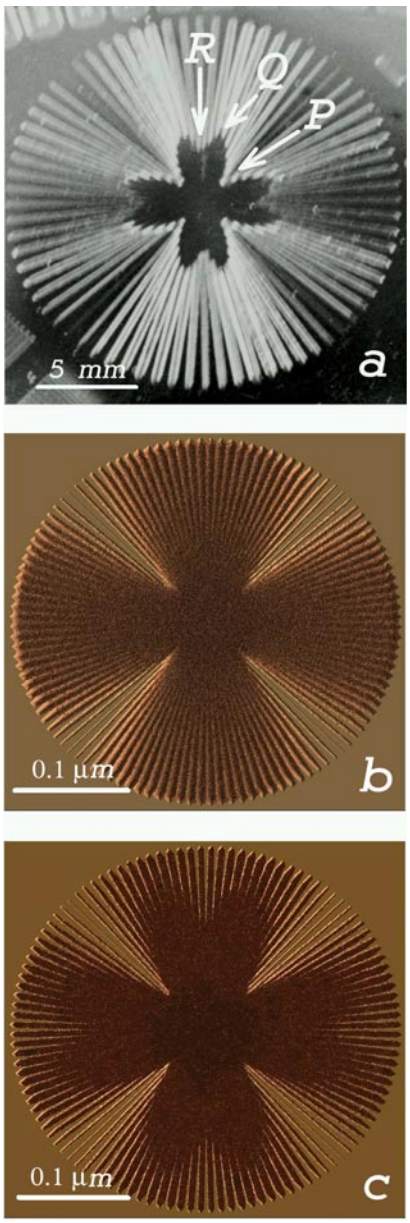

Figure 1. Etching morphology using a 'wagon-wheel' pattern on an Si(100) wafer. (a) The experimental result. (b) A combined microscopic (DFT)–macroscopic (MC) simulation, where the probabilities of bond breaking are calculated for each possible environment. (c) The results of the simulation where the indirect second-neighbour interactions have also been included. The angular distance between the spokes is 5° in the experiment and 3° in the simulations; for more details, see [22].

with high-energy electron and ion beams have shown that heavy-ion or electron irradiation induces the nucleation of small diamond crystallites inside concentric nested carbon fullerenes, 'bucky onions'.

The concentric shells may be viewed as nanoscopic pressure cells for the nucleation of diamond. Irradiating particles knock out carbon atoms from the shells, which leads to their contraction and increase of pressure towards the onion core. The interstitials catalyse sp^3 cross-links between atoms on neighbouring shells, which eventually leads to the formation of small diamond crystallites. These crystallites grow under continuing irradiation and consume the graphitic onions.

Recent computer simulations [23] have shed additional light on the atomistic nucleation and growth processes of tetrahedrally bonded carbon clusters inside fullerenes. The simulations use the density-functional-based tight-binding method [11]. This method, which contains the crucial features of quantum mechanical bonding between covalent atoms, has been successfully applied in studies of hydrocarbon molecules, carbon clusters and fullerenes as well as amorphous and crystalline carbon. The nucleation of diamond in a small, two-shell carbon onion with 300 atoms is studied with explicit molecular dynamics simulations. This low-energy allotrope is a likely core of real giant fullerenes. The simulations model the release of carbon atoms in knock-on displacements in outer shells and their transport to this core by a sequence of random atom additions with low initial velocities. The structural evolution during the injection of additional interstitial atoms is examined by a combination of molecular dynamics and quasistatic (conjugate-gradient) structural optimization.

The simulations support the general picture of diamond nucleation during irradiation. The results show that there is a critical size for arrangements of displaced atoms, below which threefold-coordinated atoms dominate but above which diamond-like-coordinated clusters can bond into completely fourfold-coordinated structures. The confinement due to the multilayer carbon fullerene appears to lower this limit and reduce the tendency of sp^3-coordinated atoms to relax back into graphitic form. The nucleation in the fullerene core involves a transition of quasi-two-dimensional curved shell structures into denser three-dimensional diamond-like structures. The simulations show that this process requires rapid energy input by incoming particles but also fast dissipation to the surroundings. One observes a transformation which can be best described as 'percolation', when initially dispersed sp^3-bonded regions fuse into one sp^3 cluster.

The simulations provide atomic-scale details of the rapid nucleation process inside carbon fullerenes under particle irradiation. On a more macroscopic level, Zaiser and Banhart [24] have presented a thermodynamical quasiequilibrium theory to explain the irradiation-induced transformation from graphite to diamond across an arbitrary interface (i.e. not necessarily a spherical graphene shell). The fact that the threshold energy for atomic displacements T_d is much larger in diamond ($\sim$30 eV) than in graphite ($\sim$15 eV) leads to the irradiation-induced reversal of phase stability, i.e. favours the growth of diamond at the expense of graphite.

5. Oxygen in silicon

Czochralski-grown ultrapure silicon is the basic material of microelectronics. It contains a small but unavoidable amount (typically 10^{18} atoms cm^{-3}) of oxygen atoms dissolved into the silicon melt from the quartz crucible during the growth process. After crystallization, this oxygen remains as a supersaturated and inhomogeneously distributed impurity in the lattice. Heat treatments are used to homogenize the oxygen distribution. The thermal treatments involve a large number of diffusion-controlled processes associated with oxygen migration, aggregation and dissociation. Oxygen complexes of varying size are known to exist. At elevated temperatures the oxygen complexes dissociate and the concentration of single interstitial oxygens (O_i) dominates in the cooled-down sample. Upon further treatment, at temperatures above 350 °C, oxygen atoms begin to diffuse and form clusters which exhibit

donor character. Initially, single O_is bind together to form interstitial dimers O_{2i}. Further migration and clustering takes place, which leads to the appearance of several families of thermal double donors (TDDs). The microscopic identification of these defects has proven quite difficult despite considerable experimental and theoretical effort.

We have recently undertaken a comprehensive computational study of oxygen-related defects and their kinetics in Si [25]. By using accurate density-functional methods we first obtain the ground-state structures for several oxygen complexes, ranging from single interstitials to clusters with more than ten oxygen atoms. For these optimal structures, we also evaluate their stable (and metastable) electronic configurations as a function of the host electron chemical potential. When the electron Fermi level moves in the semiconducting gap (due to background doping and/or temperature), the localized electronic states associated with the oxygen complex are occupied. Consequently, the complex can appear in several charge states, which have different ground-state structures. The positions of the chemical potential where the lowest-total-energy charge state changes from one to another are called *ionization levels*. Sometimes the stable charge state may change by more than one unit as a function of the chemical potential. This signals the so-called *negative-effective-U* effect, where the defect level traps two electrons rather than one; the Coulomb repulsion (U) between the localized electrons is more than compensated by the energy lowering associated with lattice relaxation around the complex. A given complex may also exhibit *metastability*, which is another consequence of the strong coupling between electronic and ionic (nuclear) degrees of freedom. A metastable defect has another (deeper) energy minimum with a different set of atomic coordinates, reachable by overcoming an activation barrier. The two states may also have different electron counts (charges) for a fixed value of the electron chemical potential: they form a bistable form of the defect complex, with different electrical activities. Transitions between the bistable minima can be triggered by e.g. photon excitation or thermal treatment, with important consequences such as persistent photoconductivity.

An important fingerprint for the various oxygen complexes is their local vibrational modes, which can be experimentally studied by infrared (IR) absorption spectroscopy. The density-functional calculations offer the possibility to calculate the local vibrational modes as well, either through the diagonalization of the dynamical matrix (within the harmonic approximation) or through the Fourier transform of the velocity autocorrelation function obtained through first-principles molecular dynamics simulation.

After establishing the accuracy of the calculations for the structures and vibrational modes of oxygen interstitials and dimers, we obtain a general picture for larger oxygen clusters, which give rise to the primary TDD family. The first three TDDs (TDD0–TDD2) consist of one four-member ring structure where two O atoms are bonded to two common Si atoms. This forms the 'core' of the TDDs, to which flanking bond-centre one or more interstitial O atoms are added. The threefold-coordinated oxygen atom is the source of the donor character. The early TDDs exhibit bistability as they have electrically neutral counterparts, the so-called X states. The calculations provide a natural explanation for the X states. They are simple chains of staggered O interstitials, without the ring structure. The higher TDDs (TDD3–TDD12) consist of similar oxygen–silicon chain–ring structures, with flanking O_is. These structural models also give a consistent explanation of the experimentally known decrease of the double-donor activation energy for the higher TDDs, as well as of their vibrational modes.

The first-principles calculations have also been extended to obtain the association, migration, restructuring and dissociation energies for a number of oxygen chains. It is remarkable that the chains can have very low effective migration energies for their concerted motion. For example, the oxygen dimer O_{2i} has a barrier of 0.98 eV when it moves via the intermediate four-membered ring (O_{2r}) configuration, to be compared against the migration

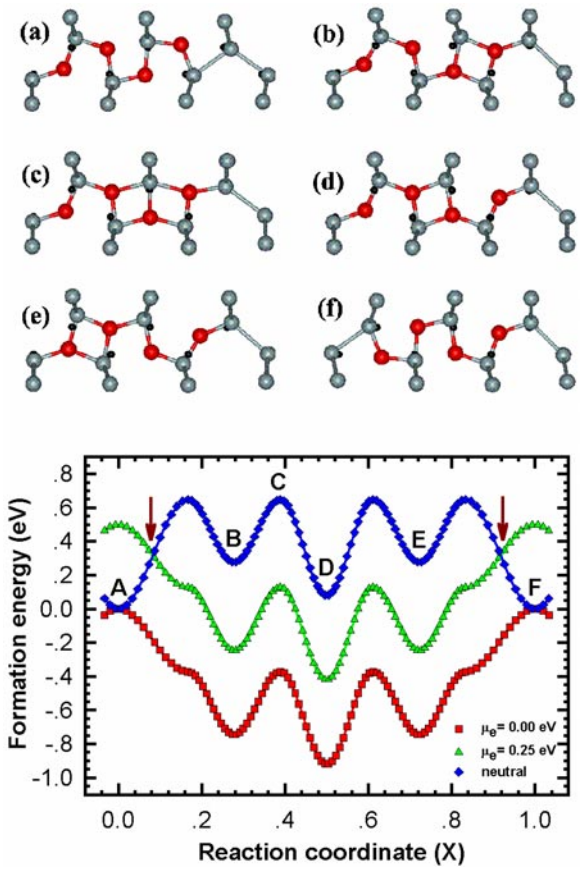

Figure 2. Structures and diffusion landscape of the four-oxygen chain O_4 along the [110] direction in crystalline Si. (a) The staggered, electrically inert configuration. (b) The TDD1 structure with electrically active, threefold-coordinated oxygen. (c) The structure of O_4 at the reconfiguration barrier. (d) The TDD2 structure. (e) The intermediate O_{2r}–O_{2i} (TDD1) configuration. (f) The final, staggered O_{4i}. The lower panel shows the total energy of the chain as a function of the reaction coordinate. A, B, C and D correspond to the structures (a), (b), (c) and (d), respectively. The top (bottom) curve is calculated for heavily n-type (p-type) materials. The middle curve corresponds to semi-insulating material. The red and blue balls denote oxygen and silicon atoms, respectively, and the black dots denote the ideal silicon crystal lattice sites. For more details, see [25].

barrier of 2.3 eV calculated for a single oxygen interstitial. Trimers and quadrimers have even lower barriers (0.6–0.7 eV). Thus not only dimers but larger oxygen complexes (the TDDs) are fast diffusers in silicon, which provides a new starting point for the explanation of the complex annealing kinetics of oxygen in crystalline Si. An example of the chain structure and diffusion path is shown in figure 2 for the four-oxygen cluster O_4.

The calculated energy parameters have been used in a general rate-equation model for the kinetics of oxygen in silicon. The model is based on the theory of diffusion-limited reactions, and includes all the possible association and dissociation reactions between oxygen complexes, as well as the isomerization and restructuring processes of single complexes. The rate equations have the form

$$\frac{\mathrm{d}[\mathrm{O}_i]}{\mathrm{d}t} = -\sum_j J_{ij} + \sum_j K_{ij} + F[\mathrm{O}_i] \tag{8}$$

where $[O_i]$ denotes the concentration of a chain with j oxygen atoms with a given ground-state structure. K_{ij} is the rate for two migrating complexes O_i and O_j to merge into O_{i+j}

$$K_{ij} = 4\pi r_0 (D_i + D_j)[O_i][O_k] \tag{9}$$

where D_j is the diffusivity of the chain and r_0 is the capture radius for the association reaction. The diffusivities are

$$D_i = D_{i0} \exp - \left(\frac{E_i^m}{k_B T} \right) \tag{10}$$

where D_{i0} is a pre-exponential factor and E_i^m the migration energy for the process in question. The dissociation rate in equation (8) is

$$J_{ij} = A r_0^{-2} (D_i + D_j) \exp - \left(\frac{E_{ij}}{k_B T} \right) [O_{i+j}] \tag{11}$$

where A is a dimensionless factor and E_{jk}^b the binding energy of the cluster O_{j+k} against dissociation into clusters O_j and O_k. Since the formation energies (and also the migration energies) of straight chains are lower than those of the less symmetric complexes, substantial isomerization of the latter into straight chains occurs. The isomerization rate is

$$F[O_i] = \nu \exp - \left(\frac{\Delta E}{k_B T} \right) [O_i] \tag{12}$$

where ν is a frequency factor and E_a the activation energy for the restructuring process. The energy parameters, including the migration barriers, depend on the value of the electron chemical potential, which in turn depends on both temperature and the number of activated donors. The temperature dependence of the processes (migration, dissociation, restructuring) is totally dominated by the Boltzmann factors with energy parameters obtained from first principles. The prefactors Ar_0^{-2}, D_{i0} and ν were obtained by fitting the concentration $[O_2]$ of dimers to the experimental [26] concentration of the 1013 cm^{-1} vibrational band at 350 and 420 °C.

Figure 3 displays calculated results for the oxygen-related TDD kinetics as compared with experimental annealing curves. Figure 3(a) shows the temperature dependence of the total formation rate of all TDDs and the corresponding loss of interstitial O_i in n-type Si. The initial oxygen concentration is $[O_i] = 1.0 \times 10^{18}$ cm^{-3}. Figure 3(b) shows the simulated annealing behaviour of the shortest oxygen chains (TDDs) in n-type silicon when the initial oxygen concentration is 8×10^{17} cm^{-3}. O_2 and O_3 denote electrically inactive (staggered-structured) oxygen chains with two and three atoms, respectively. The solid curves show the simulated results and the symbols denote experimental results [26] for local-vibrational-mode (LVM) spectroscopy of oxygen in Si. The vibrational spectra can be uniquely associated with the calculated structures and excitations. The kinetic simulations, built on microscopic calculations for the donor structures, vibrations, dissociation/association and diffusion energies, provide a consistent picture of the observed macroscopic annealing behaviour.

6. Self-diffusion in GaSb

Diffusion of host and impurity atoms in semiconductors is a crucial ingredient in several stages of the processing of device structures. For example, dopant atom profiles are sensitive to thermal treatments following implantation. The annealing influences the diffusive motion of atoms, which also depends strongly on the presence of lattice defects and impurity–defect complexes.

The recent experiments by Bracht *et al* [27] for the compound semiconductor GaSb have demonstrated that Ga atoms self-diffuse several orders of magnitude faster than Sb atoms.

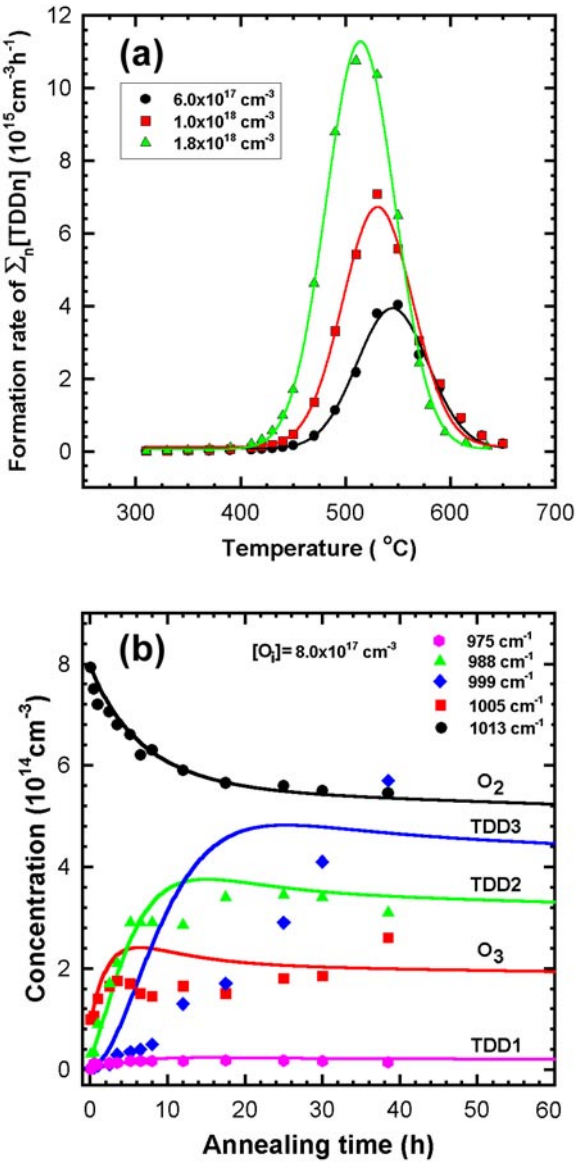

Figure 3. (a) The calculated temperature dependence of the total formation rate of TDDs and the corresponding loss rate of interstitial oxygen O_i during thermal treatment. The material is n-type Si with the initial oxygen concentration $[O_i] = 1.0 \times 10^{18}$ cm^{-3}. (b) The simulated annealing behaviour of the first oxygen chains (TDDs) in Si at 420 °C. O_2 and O_3 denote the electrically inactive chains with two and three oxygens, respectively. The initial oxygen concentration is $[O_i] = 8.0 \times 10^{17}$ cm^{-3}. The solid curves denote the simulated results and the symbols the experimental vibrational-spectroscopy data from [26]; for more details, see [25].

This result suggests that diffusion takes place rather independently in the two sublattices of the zincblende structure, and that the concentration differences of the native defects responsible for the sublattice diffusion are remarkably different. The possibility arises that the concentrations are strongly affected by amphoteric transformations between the defect types. Since defects

may possess several stable charge states, the energetics of the reactions may depend sensitively on the Fermi level position, i.e. doping of the semiconductor.

Electronic structure calculations can reveal the microscopic origin of the observed huge asymmetry in the self-diffusion coefficient. We have recently performed such calculations [28] for fully relaxed vacancies, interstitials and antisites in GaSb and estimated the native defect concentrations at typical temperatures for semiconductor growth and processing.

The structures and total energies of native defects in GaSb are calculated using DFT–LDA, as implemented through norm-conserving pseudopotentials and the plane-wave basis set. The formation energies for defects in different charge states and for different values of the material stoichiometry are calculated as

$$E_f = E_D + q(\mu_e + E_v) - n_{\mathrm{Ga}}\mu_{\mathrm{Ga}} - n_{\mathrm{Sb}}\mu_{\mathrm{Sb}} \tag{13}$$

where E_D is the total energy of the supercell containing the defect in question, q is the charge of the defect, μ_e the value of the chemical potential (Fermi level position) and E_v the energy of the valence band maximum, which also serves as the reference value for μ_e. n_{Ga} and n_{Sb} are the numbers of gallium and antimony atoms in the supercell, and μ_{Ga} and μ_{Sb} the atomic chemical potentials, respectively. The ionization levels (q/q') for a given defect are defined as the positions of the Fermi level μ_e where the charge state corresponding to the lowest energy changes from q to q' with increasing μ_e.

As-grown GaSb is always p type, irrespective of growth techniques and conditions. This is naturally explained by the calculations, which show a low formation energy (and thus a high equilibrium concentration) for the gallium antisite $\mathrm{Ga_{Sb}}$. This defect is a double acceptor. The calculated hole concentration agrees well with experimental observations. In Ga-rich growth conditions the antisite $\mathrm{Ga_{Sb}}$ concentration pulls the Fermi level down from its intrinsic value. For Ga-rich material, the other important native defect is the Ga interstitial Ga_i, which is positively charged.

For Sb-rich conditions the important defects are the gallium vacancy V_{Ga} and the antimony antisite $\mathrm{Sb_{Ga}}$. The gallium vacancy is a triple acceptor and is abundant. The anion antisite $\mathrm{Sb_{Ga}}$ has a metastable configuration when the Sb atom is displaced along the [111] direction, quite similarly to the metastability of several anion antisite defects in GaAs. However, there are no electrically active levels associated with $\mathrm{Sb_{Ga}}$. This is in important contrast to the GaAs, where the anion antisite $\mathrm{As_{Ga}}$ with its deep donor levels is responsible for the Fermi level pinning in the middle of the gap, as the microscopic model for the 'EL2' defect responsible for the semi-insulating property of as-grown GaAs. The arsenic antisite is also an important defect in GaAs thin films grown at low temperature [29].

The calculations also provide an explanation of the observed self-diffusion asymmetry. The vacancy mechanism for self-diffusion in a compound material consists of successive atomic movements to the vacant site, either from the nearest-neighbour or next-nearest-neighbour position. The relevant nearest-neighbour steps are

$$V_{\mathrm{Ga}} \leftrightarrow V_{\mathrm{Sb}}\mathrm{Sb_{Ga}} \tag{14}$$

$$V_{\mathrm{Sb}} \leftrightarrow V_{\mathrm{Ga}}\mathrm{Ga_{Sb}}. \tag{15}$$

In the first reaction, V_{Ga} can exist in negative charge states while for the complex $V_{\mathrm{Sb}}\mathrm{Sb_{Ga}}$ only the positive (1+) charge state is stable (and the neutral state metastable). The reaction (14) is therefore endothermic and requires electron transfer. The second reaction (15) is, by contrast, exothermic for all Fermi level positions. While the complex $V_{\mathrm{Ga}}\mathrm{Ga_{Sb}}$ is stable (unlike its counterpart in GaAs) and thus a possible diffusion intermediate, the strong asymmetry between the reaction energies for equations (14) and (15) suggests the nearest-neighbour mechanism.

In the nearest-neighbour mechanism the Ga and Sb atoms diffuse independently of each other via either vacancies or interstitials. The diffusion coefficient for either atom can be written with the help of the concentrations C_V and C_I (vacancies and interstitials) of the defects mediating the diffusion

$$D_{self} = d_V C_V + d_I C_I. \tag{16}$$

The diffusivities $d_{V,I}$ contain the factors related to migration (the Boltzmann factor for the migration barrier and the entropic prefactor). The concentrations can be estimated from the total energies for defects in various charge states and under different stoichiometries. The results show a large concentration of both Ga vacancies and interstitials, while the concentration of Sb vacancies is expected to be very low. V_{Sb} undergoes an amphoteric transformation to a gallium vacancy–gallium antisite complex. The complex may dissociate (more easily in n-type material) and supply extra gallium vacancies and antisites. Thus the observation of Bracht *et al* [27] for Ga-dominated diffusion can be understood by the significant concentration differences of native defects in the two sublattices. However, it seems that the diffusion may not be described by the (sublattice) vacancy mechanism alone, since the relative concentrations of interstitials are also non-negligible.

7. Summary and conclusions

This paper has discussed some current trends in computational materials science. This field is undergoing a phase of rapid progress, as increasingly complex problems associated with real materials and their processing can be addressed. This progress naturally builds on the solid foundation of condensed matter physics, especially as regards the atom-scale phenomena. While the conceptual basis for many physico-chemical properties is well understood, condensed-matter research continues to provide surprises and new theoretical challenges in its less charted areas. Examples of the latter include exotic superconductors and correlated electronic phases in low-dimensional systems. However, there is ample reason to believe that the gamut of materials phenomena which succumb to quantitative modelling by numerical techniques will continue to grow.

In computational materials science, powerful techniques such those based on DFT and on molecular-dynamics simulations are now used to attack problems which most researchers would not have even considered just a decade ago. These include free-energy calculations at finite temperatures, noncollinear magnetic structures, extended defects such as dislocations and grain boundaries, and even systems far away from equilibrium. The modelling is not only qualitative but also quantitative. This is important, as for many scientifically and technologically important questions in materials, 'the devil is in the detail'.

The progress naturally owes a lot to the rapidly improved access to unprecedented computing power at affordable costs. Materials modelling has been one of the scientific beneficiaries of Moore's law, and it is perhaps gratifying to think that advances in computational materials science are paying back to microelectronics and computer engineering, as new materials and structures help to design faster and cheaper processors and memories.

Another important contributor to the progress is new algorithms and computational methods applied in solving the relevant equations, ranging from iterative matrix diagonalizations to novel MC tricks. Many of these algorithms draw their inspiration from the physical world. Simulated annealing, genetic algorithms and lattice-gas automata are typical examples of such algorithms, which also find many applications outside physical sciences. Again, one can think of this as a payback from computational physics to 'physical computing'.

A number of examples of materials simulation have been briefly described in this article. These examples, which have been chose to illustrate the multiscale, multifaceted nature of the field, come from the work carried out at the Laboratory of Physics at HUT. They constitute just a tiny and not necessarily representative sample of the widespread activities in computational materials science. The interested reader is referred to the rapidly growing scientific literature as well as to descriptions of large research initiatives in the area.

Acknowledgments

This work has been done in collaboration with Roope Astala, Juhani von Boehm, Adam Foster, Miguel Gosalvez, Mikko Hakala, Markus Kaukonen, Young Joo Lee and Martti Puska. The research at COMP is supported by the Academy of Finland through the Centres of Excellence Programme (2000–2005).

References

[1] Broughton J, Abraham F F, Bernstein N and Kaxiras E 1999 *Phys. Rev.* B **60** 2391
Ogata S, Lidorikis E, Shijomo F, Nakano A, Vashishta P and Kalia R K 2001 *Comput. Phys. Commun.* **138** 143
[2] Shenoy V B, Miller R, Tadmor E B, Phillips R and Ortiz M 1998 *Phys. Rev. Lett.* **80** 742
Garcia A L, Bell J B, Crutchfield W Y and Alder B J 1999 *J. Comput. Phys.* **154** 134
[3] Dreizler R M and Gross E K U 1998 *Density Functional Theory* (Berlin: Springer)
Nieminen R M 1999 *Curr. Opin. Solid State Mater. Sci.* **4** 493
[4] Williamson A J, Hood R Q and Grossman J C 2001 *Phys. Rev. Lett.* **87** 246406
[5] Onida G, Reining L and Rubio A *Rev. Mod. Phys.* at press
[6] Beck T L 2000 *Rev. Mod. Phys.* **72** 1041
Chelikowsky J R, Saad Y, Ögut S, Vasiliev I and Stathopoulos A 2000 *Phys. Status Solidi* **217** 173
Heiskanen M, Torsti T, Puska M J and Nieminen R M 2001 *Phys. Rev.* B **63** 245106
[7] Baroni S, de Gironcoli S and Dal Corso A 2001 *Rev. Mod. Phys.* **73** 515
[8] Frenkel D and Smit B 1996 *Understanding Molecular Simulation* (London: Academic)
Rapaport D 1995 *The Art of Molecular Dynamics Simulation* (Cambridge: Cambridge University Press)
[9] Sorensen M R and Voter A F 2000 *J. Chem. Phys.* **112** 9599
[10] Car R and Parrinello M 1985 *Phys. Rev. Lett.* **55** 2471
[11] Frauenheim Th, Seifert G, Elsner M, Hajnal Z, Jungnickel G, Porezag D, Suhai S and Scholz R 2000 *Phys. Status Solidi* **217** 41
Bowler D and Gillan M 1998 *Comput. Phys. Commun.* **112** 103
Ordejon P 1998 *Comput. Mater. Sci.* **12** 157
[12] Brenner D W 2000 *Phys. Status Solidi* **217** 23
[13] Landau D P and Binder K 2000 *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge: Cambridge University Press)
[14] Jonsson H, Mills G and Jacobsen K W 1998 *Classical and Quantum Dynamics in Condensed Phase Simulations* ed B J Berne, G Ciccotti and D F Coker (Singapore: World Scientific)
[15] Kaukonen M, Sitch P K, Jungnickel G, Nieminen R M, Pöykkö S, Porezag D and Frauenheim Th 1998 *Phys. Rev.* B **57** 9965
[16] Geissler P L, Dellago C, Chandler D, Hutter J and Parrinello M 2001 *Science* **291** 2121
[17] Bortz A B, Kalos M H and Lebowitz J L 1975 *J. Comput. Phys.* **17** 10
[18] Chopard B and Droz M 1998 *Cellular Automata Modeling of Physical Systems* (Cambridge: Cambridge University Press)
[19] Dhatt G and Touzot G 1984 *The Finite Element Method Displayed* (New York: Wiley)
[20] Rudd R E and Broughton J Q 2000 *Phys. Status Solidi* **217** 251
Lundqvist B I, Bogicevic A, Dudiy S, Hyldgaard P, Ovesson S, Ruberto C, Schröder E and Wahnström G *Comput. Mater. Sci.* at press
[21] Waite T R 1957 *Phys. Rev.* **107** 463

[22] Gosalvez M A, Nieminen R M, Kilpinen P, Haimi E and Lindroos V 2001 *Appl. Surf. Sci.* **178** 7
Gosalvez M A, Foster A S and Nieminen R M *Phys. Rev. Lett.* submitted
[23] Astala R, Kaukonen M, Nieminen R M, Jungnickel G and Frauenheim Th 2001 *Phys. Rev.* B **63**
[24] Zaiser M and Banhart F 1997 *Phys. Rev. Lett.* **79** 3680
[25] Young Joo Lee, von Boehm J, Pesola M and Nieminen R M 2001 *Phys. Rev. Lett.* **86** 3060
Young Joo Lee, von Boehm J and Nieminen R M 2001 *Appl. Phys. Lett.* **79** 1453
Young Joo Lee, von Boehm J, Pesola M and Nieminen R M *Phys. Rev.* B at press
[26] Åberg D, Svensson B G, Hallberg T and Lindström J L 1998 *Phys. Rev.* B **58** 12 944
[27] Bracht H, Nicols S P, Walukiewicz W, Silveira J P, Briones F and Haller E E 2000 *Nature* **408** 69
[28] Hakala M, Puska M J and Nieminen R M *J. Appl. Phys.* at press
[29] Staab T E M, Nieminen R M, Gebauer J, Krause-Rehberg R, Luysberg M, Haugk M and Frauenheim Th 2001 *Phys. Rev. Lett.* **87** 045504