Modelling and simulation of solids
Editorial overview
Marshall Stoneham* and Sokrates T Pantelides[†]

Addresses
*Physics Department, University College London, Gower Street, London WC1E 6BT, UK; e-mail: a.stoneham@ucl.ac.uk
†Department of Physics and Astronomy, Vanderbilt University, Tennessee 37235, USA; e-mail: pantelides@vanderbilt.edu

The quest for a quantitative description of the microscopic properties of solids has its roots in the 1930s, soon after the development of quantum mechanics. Pioneers like Mott, Seitz and Hume–Rothery asked how far one could go from the most fundamental descriptions of atoms to achieve an understanding of bulk solids and their properties. The Hartree and Hartree–Fock approaches to the many-electron problem made it clear that often one could work with an effective one-electron potential acting on each electron. Bloch's theorem set the stage for calculating the energy bands of perfect solids, and the understanding of the differences between insulators, semiconductors, and metals. New ideas emerged, such as Landau's recognition of self-trapping, Mott and Littleton's approach to the large polarisation energies associated with charged defects or impurities, and the emerging recognition of dislocations and their roles.

As regards electronic structure, physicists pursued several distinct lines of research which were rewarded by substantial successes from the 1960s and 1970s. The issue of how to construct the effective one-electron potential led to the pseudopotential and related developments. At the same time, quantum chemists were making Hartree–Fock methods much more accessible, with systematic approximations for more complicated molecules. Notable among these methods were the Zero Differential Overlap methods, which involve systematic approximation from full Hartree–Fock methods (or, as was shown later, from density functional methods [1]), plus some empirical information.

The pseudopotential methods and the quantum chemical methods gained rapid exploitation in studies of defects in solids, where (especially in ionic systems) they provided both a framework for understanding and even accurate descriptions of both ground and excited states of defects, of impurities, and of self-trapped states. The main experimental emphasis was spectroscopy, and this made available extremely wide-ranging and detailed data for certain simple defects. Such key defects ranged from the F centre (an electron in an anion vacancy) and the self-trapped hole in alkali halides to the vacancy in diamond. Theory gave remarkably good answers to some, but not all, the ques-tions, even including lattice distortion and polarisation, accurately [2].

In the late 1970s a new focus for electronic structure calculations emerged as advances in vacuum science and experimental techniques provided systematic data on clean surfaces. One innovation was the use of supercells (large cells of atoms repeated periodically) which streamlined calculations [3]. For point defects in semiconductors, Green's functions provided a breakthrough in computing discrete energy levels in the band gap [4,5]. By the late 1970s advances in experiment and theory were demanding improved calculations of total energies in order to determine lattice relaxations and reconstructions. At about the same time, accurate calculations performed on bigger computers led to the realisation that density functional theory with the Kohn–Sham local density approximation for exchange and correlation, the best prescription for constructing an effective one-electron crystal potential, yielded poor band gaps. As a result of this difficulty, for the last 15 years or so, total-energy calculations have dominated the electronic structure scene. In parallel the 'band gap problem' has been solved by invoking an old theory by Hedon and Lundqvist [6]. The method, however, is computationally very demanding, and is not widely used except in the simplest systems such as perfect crystals and other cases involving small cells of atoms More recently, an even more ambitious undertaking is under way: the solution of the many-electron problem, exactly, by Monte Carlo sampling methods. The results so far are promising and exciting (see the review by Lubos Mitas, pp 696–700), but the computational demands are such that practical and routine use of this method will take some time. Even with this promising approach, it will be hard to find a means to handle complex dynamical behaviour observed in electronic excited states

Supercells have become the primary method of choice for total-energy calculations in semiconductors, although other methods, such as embedding, are developing rapidly, and are favoured for ionic systems. A recent major development was the introduction of the Car–Parrinello method [7], which showed that the electronic and ionic degrees of freedom of an assembly of atoms can be treated on an equal footing. Car and Parrinello chose the context of the Hohenberg–Kohn density functional theory, but the combination of electronic structure and classical nuclear dynamics has since taken diverse forms.

Classical molecular dynamics, in which Newton's equations are solved numerically for particles interacting by assumed interatomic potentials, has been used for half a century. The original idea was to describe bonding in mol-

ecules and solids by expanding the total energy as a sum of two-body (pair), three-body, and higher-order interatomic potentials. In the 1960s and 1970s, suitable computer algorithms were developed to handle the time-evolution problem while controlling external variables, such as pressure, temperature, and so on. In the case of ionic systems, and especially in approaches using the shell model, the accuracy of the dynamic methods and the corresponding static methods proved remarkably good. For semiconductors and metals, the situation was far less satisfactory, and there have been tremendous efforts in constructing ever more sophisticated interatomic potentials. Car–Parrinello dynamics eliminated the interatomic potential by exploiting the quantum mechanics of the electronic system directly. Speed is still a major factor: classical molecular dynamics can handle millions of atoms in a supercell, where Car–Parrinello dynamics can only handle hundreds. Some of the ways round this difficulty involve embedding a Car–Parrinello central core in a classical environment.

What we have described so far starts from the smallest scale (length scale or time scale) and from the most basic methods. But needs are more wide-ranging, especially when technology is involved. There are also innumerable variants of the methods we have described, hybrid methods, and semiempirical approaches. These variants are often used by scientists who need to understand some significant, often complex, practical problem. They may wish to decide which processes are important before using a more fundamental approach; they may need to look at non-equilibrium situations, or excited states, or long time scales, or indeed any of the situations which are hard to handle from first principles.

Modelling and simulation are becoming increasingly ambitious. On the one hand, there are calculations which aim at an increasingly precise and detailed description. Assumptions about exchange and correlation are minimised, and empirical elements are replaced wherever possible. In the best cases, the improvements in detail yield improvements in accuracy, and provide the potential to tackle problems previously intractable. On the other hand, there are ambitions to tackle increasingly large systems. These have different challenges. The methods which are best for studies concentrating on a few atoms may be inefficient for larger structures; there may be more alternative structures possible than are obvious from the smaller scale systems; the interpretation itself may be difficult without other aids. There is therefore a need to link experience at the macroscopic level with the understanding from an atomic scale. The present issue addresses both of these ambitions. The review on 'Quantum Monte Carlo' methods (by Lubos Mitas) describes the approaches closest to first principles currently in use in modelling. The review on ferroelectrics (David Vanderbilt, pp 701–705) shows how the problems in the atomistic modelling of these tricky systems can be solved. Not long ago, even the static dielectric constant of simple polar dielectrics presented

problems; now the far harder problems of ferroelectrics are being treated with great success.

The other articles all address the large length scales, but illustrate a variety of approaches. The discussion by Jerzy Bernholc et al. (pp 706–715) on nanotubes shows how strikingly state-of-the-art modelling can work alongside novel experiments in the understanding of a new class of materials. In these tube-like structures, it proves helpful to develop analogies with macroscopic tubes, as a convenient way of displaying calculated properties. This is one of the several classes of link between microscopic and macroscopic descriptions included in this issue. Catherine Priester and Michel Lannoo's (pp 716–721) study of quantum dots brings in macroscopic concepts like surface tension on the mesoscale. The study of 'Semiconductor strained layers' by Suresh Jain and co-workers (pp 722–727) emphasises elasticity theory in the discussion of metastable strain-layer systems and the dislocation engineering which underpins their technological promise. John H Harding (pp 728–732) looks more broadly at mesoscopic modelling. His work shows that one can go a long way to achieving 'whole process modelling' of plasma-sprayed coatings, covering the heat transfer in the plasma, the dynamics of the splat, the mesostructure of the thermal barrier, and the effective heat transport and macroscopic elastic constants of the coat.

The generic problem is the linkage of atomic-scale information to the mesoscopic scales of complex materials and the macroscopic continuum (engineering scales). If the importance of macroscopic physics remains clear, the nature of the link to atomic-scale physics is neither unique nor always clear. Mesoscopic modelling takes various forms. John H Harding notes the differences between those defined from below, moving from atoms to larger units, and compares them with those defined from above, moving from the larger scales to the smaller. There are other important differences. Perhaps the major one is between those systems for which an averaging is sufficient, and those for which one has to construct many realisations of the microstructure. Averaging is often done by an effective medium approach, and is suitable for predicting properties such as density, or for dielectric constants or conductivities in suitable cases. However, fracture, electrical breakdown, or superconductor critical currents, will depend on specific details in the microstructure; extremal statistics may be appropriate, and averaging is not acceptable. In such cases, one must check the specific observable property for many realisations of the microstructure. There are further differences in methods, of course, especially when there are time-dependent phenomena.

Of course, for larger systems, there are other issues as well, as noted above: fracture and crack propagation, dielectric breakdown, and other such phenomena are acutely sensitive to boundary conditions and to structural detail. In such cases one is trying to sample special extreme conditions, and special approaches must be used. Yet the solutions of

difficult problems like these are typical of some of those which bring physics into direct contact with engineers. And, if we are going to make contact with engineers in a meaningful way, this means that those who do the modelling and simulation should be able to predict what these engineers see, or need for their work. Sometimes this is a matter of good graphics (a serious component of computational fluid dynamics); sometimes it is a matter of code compatibility; and sometimes it is a matter of developing the right sorts of property packages. There is an interesting example cited by John H Harding, in which many realisations are needed to establish the microstructure of a thermal barrier layer, yet where an effective medium theory is sufficient to predict the properties of that coating.

John H Harding also makes the point that mesoscopic models can be top-down or bottom-up. The survey by Jerzy Bernholc *et al.* is an example of bottom-up, starting from a modest number of atoms, but continuing these atomistic calculations to much larger systems. Examples of top-down calculations are covered in the discussion on ceramic superconductors by Tuck C Choy (pp 733–738) and the reviews by Suresh Jain *et al.* and by Catherine Priester and Michel Lannoo. Both these papers show how the range of semiconductor materials can be extended, in the one case by exploiting strain, in the other case by exploiting size and surface energies. In most of the top-down theories, it is sufficient to use averaged properties, suitably generalised as to specific features. But the issue of many realisations also occurs for systems with relatively few atoms, such as the quantum dots discussed by Catherine Priester and Michel Lannoo, or the strain layers described by Suresh Jain and his co-authors. In these systems, alloy fluctuations can be important. In layered systems, these can localise excitons; in quantum dots, there will be dot-to-dot variations in critical properties.

If there are no signs of *ab initio* calculations on lumps of matter at the human scale, there are certainly useful calculations at all scales from atomic sizes (0.1 nm) to large engineered structures (1 km), some 13 orders of magnitude. The time scales of interest present even greater problems, stretching from cascades in radiation damage (1 fs) to thousands of years (or even longer, if geological materials are included), at least 24 orders of magnitude. The articles on mesoscopic modelling aim to cover the length scales. There is no comparable theory to cover the wide range of times. Certainly, there are tools (including diffusion theory) which are the basis of many studies of long time scales. Yet the link between the short effective times of molecular dynamics and the long times of many important processes has still be completed. There are situations in which events on very short time scales determine behaviour at a far later stage, and there limits on how well we can understand these links. As we observed in a previous introduction, in the radiation damage of steels, events in the first few femtoseconds affect the development of microstructures which influence ductility after tens of years (24 orders of magnitude in time). This

ductility may decide whether a power reactor may be run for a further ten years, which would make an enormous difference to reactor economics. Human scales and values define the needs for materials; whether these desires can be met depends on the atomic scale and on understanding.

Collectively, the approaches show how modelling can have impact, crucial at several different levels. Simple theories offer a framework for experiment and for applications. Modest approaches (less than the state-of-the-art) can identify the critical areas. At best, some observable quantities can be predicted more accurately than measured.

What is clear is the importance of macroscopic methods: elasticity theory, continuum electromagnetic theory, and their generalisations. Obviously, these are very simple approaches if applied at the atomic scale. But the continuum theories provide a limit for atomic treatments of regions of larger and larger sizes, and they provide a tool for embedding sophisticated treatments of a small subsystem within a representation of its real environment.

What are the remaining challenges? The need for ways to handle highly non-equilibrium excited states is still barely addressed. For some systems, as in radiation damage, when the electrons do not have a Fermi distribution and the degree of excitation varies rapidly in space. There is still no routine way to treat non-adiabatic transitions, especially those between excited states; here, quantum nuclear dynamics is needed, as well as electron dynamics. There is still the challenge of designer materials. It is not sufficient to choose a material, or decide a desirable material structure. There has to be way to make it. The excitement of oxide superconductors would have been even greater if ideas for better materials had been followed by ideas from modelling for ways to make these materials. At the mesoscopic scale, composites are a means to enhance properties, and composite design is a reality. At the atomic scale, organic materials have been designed successfully. Semiconductor systems have made a lot of progress. Other inorganic materials are still at an early stage, but promise a new meaning for the words 'molecular engineering'.

References

1. Lindholm E, Lundqvist S: **Semiempirical MO methods, deduced from density functional theory.** *Phys Scr* 1985, **32**:220.

2. Stoneham AM: *Theory of Defects in Solids.* Oxford: Oxford University Press; 1975.

3. Cohen ML, Schlüter M, Chelikowsky JR, Louie SG: **Self-consistent pseudo-potential method for localized configurations: molecules.** *Phys Rev B* 1975, **12**:5575.

4. Baraff GA, Schlüter M: **Self-consistent Green's functions calculation of the ideal Si vacancy.** *Phys Rev Lett* 1978, **41**:892.

5. Bernholc J, Lipari NO, Pantelides ST: **Self-consistent method for point defects in semiconductors: application to the vacancy in silicon.** *Phys Rev Lett* 1978, **41**:895.

6. Hedin L, Lundqvist S: **Effects of electron–electron and electron–phonon interactions on the one-electron states of solids.** *Solid State Phys* 1969, **23**:1.

7. Car R, Parrinello M: **Unified approach for molecular dynamics and density functional theory.** *Phys Rev Lett* 1985, **55**:2471.