Defects in semiconductors and oxides: where are the gaps in first principles theory?

Marshall Stoneham

Centre for Materials Research, and London Centre for Nanotechnology, Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

Abstract

Defects in semiconductors and oxides have been the subject of some of the most sophisticated approaches to modelling and simulation. The powerful, widely used methods can give the impression that all technologically important materials problems can be addressed reliably. But is this so? This paper looks at some of the gaps in first principles theories and at the situations that still warrant attention. Excited states, non-equilibrium systems and non-adiabatic transitions, the correct handling of different length and time scales and the prediction of characteristically quantum behaviour all present challenges. Whilst the emphasis is on semiconductor and oxide systems, the wider context of materials science points to further issues that should not be overlooked.

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

1. Introduction

Defects of semiconductors and oxides are not just diverse, but they are also seriously important. They control the value of gemstones, the stabilities of materials for radioactive waste containment and the functionality of most materials for the information technologies. It was the semiconductor industry and the nuclear industry that developed the first general purpose codes for defect modelling, and these codes grew in power alongside the research and development projects that needed them. Strategies for modelling and simulation evolved alongside hardware and software. It is no surprise that oxides and semiconductors featured strongly in leading edge developments.

Nearly half a century on, one gets a vivid picture of the enormous vigour and power of codes for materials modelling. Many calculations assert 'first principles' or '*ab initio*' status, though there are often significant working approximations that make one wonder how appropriate such descriptions are. But, even if these descriptors were correct, should we believe that we can answer all the important questions materials technology might pose?

I believe it is timely to try to identify what significant problems there are that we still do not know how to do. Some of the gaps will depend on new hardware and software, especially

when modelling quantum systems. Some gaps will be recognized only after some social or technological change has brought them into focus. Among the developments likely to stimulate innovation could be novel nanoelectronics or the fields where physics meets biology. Yet even traditional materials studies may involve features, such as other length scales, that we cannot reach with methods that are bottom-up, starting from atomic descriptions (Stoneham and Harding 2003, Stoneham *et al* 2001). Ice is a perfectly good insulator, and in many ways a simple oxide, but we may be a long way from a first principles study of the 'mouth feel' of ice cream, despite its commercial significance. Some further gaps exist because, too often, we are slaves to fashion. We are easily drawn away from unpopular (otherwise translating as 'too difficult') topics. Examples might include excited state spectroscopies or electrical breakdown.

The existence of gaps should not overshadow the immense progress in mainstream solid state physics. Much successful pioneering work on materials modelling was based on very simple potentials and non-self-consistent electronic structure. Today, potentials are sophisticated, respectably accurate over a wide range of situations. Self-consistent electronic structure for molecular dynamics is routine for adiabatic energy surfaces. Codes are robust, and there is awareness of problem areas. This all gives us confidence. But there are hints that any problem can be solved in a way that can be called 'first principles.' Is that true? Can almost any process be modelled with success and without major approximations? Sadly, this seems highly unlikely. There are at least three groups of challenges.

The first group of challenges concerns electronic excited states and other non-equilibrium systems. 'Non-equilibrium' encompasses many ideas. Predicting some equilibrium states, such as the crystal structure of lowest energy, is only rarely the major issue in materials modelling or its applications. For many systems, the exploitable characteristics occur in metastable forms (e.g. most steels, diamond), not the state of lowest free energy. Predicting which crystal structure will result from a particular growth process can be very hard indeed, especially for larger organic molecules. Getting a right structure experimentally (say, growing diamond instead of graphite) is serious, of course, both harder and more serious than a computer prediction. Which structure you obtain in practice can be crucial in performance, e.g. left- and right-handed molecules may be the same in energy but have utterly different biological consequences. The best simulations will recognize that process is often more important than structure.

A second set of challenges concerns timescales, from femtoseconds to tens of years, and length scales and the links between microscopic (atomistic) and mesoscopic (microstructural) scales for hierarchical phenomena. Will radiation damage events (timescale picoseconds) cause embrittlement in a weld in a nuclear reactor (timescale 30 years)? It is far from obvious that the usual 'spanning the length scales' description encourages the most appropriate strategies. How do organized structures emerge, especially when soft materials appear to control hard ones (e.g. Stoneham 2007)? Protein folding and replication of DNA do not represent materials fabrication in its usual sense, but the physics and chemistry of the way life forms are created will surely include phenomena that materials modelling should be able to address. There are issues of strategy here, and I, for one, doubt the effectiveness and accuracy of huge codes covering all length and time scales at the same time. The full range of behaviours must be embraced, of course, including atomic, mesoscale and human (engineering) scales. But usually it will be most effective to identify and sort out the dominating scales first (e.g. Stoneham and Harding 2003).

A third challenge lies in understanding the quantum physics of highly correlated systems, for which intuition based on the human scale is not a safe guide. Most first principles codes have some quantum electrons and classical ions and rarely calculate the entanglement that underlies

quantum information processing or even the coherences often encountered at the mesoscale. Whether we can usefully model all interesting quantum situations without quantum computing capability seems doubtful, though ingenuity may take us quite a long way.

These three challenges seem to demand new approaches. There are, of course, important practical issues at a more incremental level, cases where one would expect current methods to manage well, yet they have not shown the success one would hope for. Even for the well-known substitutional donors in silicon, there are few predictions of valley-orbit splitting, and—to my knowledge—no first principles predictions of the unusual trends in ionization energies and ground state hyperfine constants along the column of Group V donors (P, As, Sb, Bi). There are, of course, plenty of empirical interpretations, many reviewed in Stoneham (1975). The diffusion of O interstitials in silicon shows a worrying discrepancy between experiment and a number of careful calculations by experienced workers. This experiment (Watkins *et al* 1982) shows an activation energy ε_{exp} of about 2.5 eV over many orders of magnitude in rate, from the lowest temperatures upwards. The calculations give a much *lower* value $\varepsilon_{\text{theory}}$, typically about 1.5 eV. If $\varepsilon_{\text{theory}}$ were bigger than ε_{exp}, it would be natural to assume some diffusion channel had not been identified. But having ε_{exp} bigger than $\varepsilon_{\text{theory}}$ is hard to explain. It is not possible to argue that really $\varepsilon_{\text{theory}}$ is correctly the minimum saddle point energy and that kinetic reasons make a diffusion path over a higher barrier more effective, partly because the pre-exponential factor in the diffusion rate is especially high (inconsistent with any kinetic suppression) and partly because the value of ε_{exp} holds down to the lowest temperatures, where the Arrhenius factor would surely overwhelm any kinetic limits.

2. Special excited states

Why are excited states important, especially in semiconductors and oxides? First, and most obviously, excitation energies and excited state symmetries both label an impurity ('the 588 nm defect') and characterize it. A calculation that merely produces a ground state charge density has little value until it can connect with what can be measured. Secondly, excited electronic states may differ from the ground state in symmetry, spin or spatial extent. So one can control these properties of a defect or impurity by appropriate sequences of electromagnetic pulses. When the excited states are more sensitive to applied fields, as they often are, there is extra control. This control can be used in several ways, perhaps for laser or maser action, or to control quantum entanglement. For effective control, a defect may need special properties, e.g. electron–phonon coupling may be crucial (characterized by the Huang–Rhys factor, cf Stoneham (1975, chapters 10 and 14)) and determines whether one gets sharp or narrow lines or a fast non-radiative transition rate. Even in the most conventional microelectronic devices, a common initial step in carrier capture process is via an excited state. If there is substantial electronic excitation, as when a fast heavy ion passes by, the carriers can store energy and redistribute energy, and this controls many subsequent damage processes.

Excited states offer new ways to control systems, especially at the nanoscale, where traditional methods based on thermal diffusion are somewhat indiscriminate. What electronic excitation enables can range from materials modification to quantum coherence (Itoh and Stoneham 2001). Underlying these roles of excited states are some general classes of phenomena (figure 1). Thus, in materials modification, a crucial component relies on energy *localization*, whether ionic or electronic. If atoms are to be displaced, some minimum energy is needed, perhaps to break a bond in some sense. This energy must be associated with specific *local* atomic motions. Often the key is self-trapping, the immobilization of excitations (excitons or electron or holes separately) by local lattice deformation.
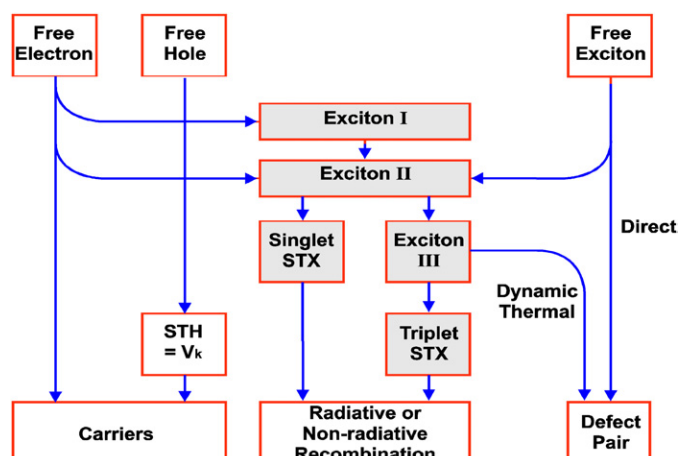
Figure 1. (Itoh and Stoneham (2001), reprinted with permission). Competing channels in the recovery of ionic solids excited above their band gaps. The diversity of behaviour following excitation, including radiation damage, should not be oversimplified.

The amount of energy available for a process is crucial, yet many current methods are disappointing in predicting available energies, whether trapping or self-trapping energies. This can be because of the very different natures of the ground and excited states, especially if the excited state is delocalized. Often, electronic excited states have features qualitatively different from the ground state, notably their equilibrium geometry and their degeneracy or near-degeneracy. Perhaps the two states differ greatly in their sensitivity to electron correlation or to self-interactions. Still more tricky is understanding the many ways that the excited state's excess energy can be dissipated by radiative or non-radiative transitions.

Charge localization is a key process that offers one simple means of guiding energy localization. But again one is often concerned with an energy difference between a localized and a delocalized state, and this may be hard to calculate because the working assumptions may differ. Likewise, current methods for predicting self-trapping may disappoint: the local density approximation (LDA) often fails to predict it, even when experiment is unambiguous, and Hartree–Fock (HF) methods predict localization when it should not. Whether or not self-trapping (or trapping) occurs has especially significant effects on carrier mobilities. When there is incoherent hopping of a small polaron, the mobility will be less than $0.1\text{--}1\,\mathrm{cm^2\,V^{-1}s^{-1}}$, limited by one hop per lattice vibration period. For large polaron behaviour, the mobility will exceed $1\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ or thereabouts, the lower limit corresponding to a mean free path of the order of interatomic spacings. Polaron behaviour is remarkably diverse, and a crucial component of the striking behaviours of oxides, including colossal magnetoresistance, room temperature ferromagnetism of sp oxides and probably high temperature superconductivity.

Charge transfer underlies the build-up of space charge. Static charge build-up can have technological consequences but, for modelling, is important mainly for indirect reasons, e.g. because a macroscopic field can develop that influences subsequent events, such as radiation damage or laser ablation. The nature of charge transfer is often obscure. It is standard to refer to Förster–Dexter mechanisms or to assume that some form of 'tunnelling' is involved, but the details are rarely checked. Tunnelling has naturally difficult features, since rates often depend exponentially on features that are not known well. With care, simpler systems with high symmetry can be solved, e.g. H tunnelling within the NVH$^-$ centre in diamond

(Shaw *et al* 2005). However, much harder are low symmetry cases, such as proton tunnelling in biological enzyme reactions (van der Kamp *et al* 2008, Johannissen *et al* 2008) or as a check of early proposals for cancer initiation (Löwdin 1963) or electron tunnelling between semiconducting organic polymer chains in a disordered network. Whilst quantum tunnelling is a common phenomenon, its modelling is often primitive, intended more to make use of simple analytical models than to represent the system accurately. Such models do have considerable interpretive value, of course, as in studies of scanning probe microscopy.

Energy transfer is a third process involving an excited state. Light emission or atomic displacements do not need to occur at the ionic site originally excited, but can occur at more distant sites following intermediate steps. Thus excitation of a disordered semiconducting polymer system may be followed by the excitation moving to another site where the excited state energy is lower; this leads to a red shift in emission. Photosynthesis includes an energy transfer step, with some evidence of quantum coherence (Engel *et al* 2007). But simple motion of excited conduction electrons can lead to energy transport. When the electrons have been excited by the passage of high energy heavy ions in metals, the extent of energy transport can determine the extent of residual damage (Stoneham 1989) and whether or not an ion track will be seen (Duffy *et al* 2008).

Fourthly, there is energy storage. Energy sinks can delay damage processes, and sometimes change their character. In the case of ion tracks, just mentioned, even conduction electrons in a metal can store energy for long enough to affect outcomes. Often energy storage occurs in combination with, say, charge build-up, as in electrical breakdown. This can lead to puzzles. In electrical breakdown, for instance, the average energy per atom stored at breakdown is less than 0.5 meV, far less than the 25 meV thermal energy or the few electronvolts needed to break a chemical bond. There must be some means to concentrate the energy. When there is a bandgap greater than the energy to cause irreversible change, one can imagine possible processes (see, e.g. Itoh and Stoneham (2001)) but puzzles can remain. For example, 5 eV light incident on MgO, with an energy less than the bandgap, can cause emission of ions of 80 eV energy (Dickinson *et al* 1993, Webb *et al* 1993a, 1993b). How might this happen? Suppose the charge responsible for driving emissions with this energy is created by electrons moved by incident photons of energy E_{ph}. These photons must cause a charge to build up in a small region of radius R. To add another electron (or hole) into this region, the energy from that one photon must be able to overcome the Coulomb barrier from charges already in the region. The limit is a charge Q_{max} when the energy E_{ph} just fails to move an electron by distance a across the boundary of the region, i.e. $E_{ph} \sim Q_{max}[R^{-1} - (R+a)^{-1}]$, so $Q_{max} \sim R^2 E_{ph}/a$. Then the maximum energy that can be given to a particle of charge Z leaving the surface of the region is ZQ_{max}/R, or $ZE_{ph}R/a$. It seems most unlikely that R is more than a few times larger than a, there should be a limit on ion emission energy of a few times E_{ph} for particles accelerated solely by local surface charge. If this primitive analysis is correct, the factor *ZR/a* is about $80/5 = 16$, large but not outrageous. But a full first principles calculation would both check these ideas and, probably, give insight into other phenomena. There are still more puzzling cases awaiting explanation, such as the emission of x rays with tens of kiloelectron energies on peeling sticky tape (Camara *et al* 2008).

For excited states, there are serious limitations of parameter-free approaches, whether through scaling problems, as in post-Hartree–Fock theories, or because it is hard to extend ground state approaches such as density-functional theory. There are plenty of scientific and technological problems for which such parameter-free approaches do not seem able to provide answers. In some cases, much simpler methods succeed, e.g. in predicting from the optical absorption date just which of the many F centres in ionic crystals will show luminescence (see, e.g. Itoh and Stoneham, section 3.5.2). We can either pretend such difficult questions do not

exist or we can try to appreciate better the virtues of those theories that do not earn the label of '*ab initio*' but yet provide us with quantitative understanding.

3. Away from equilibrium and the steady state

Metastable states.   In a semiconductor, it is natural to assume that carrier injection will be followed by rapid processes leading to equilibrium or to some driven steady state. But when an exciton is created, or a muon implanted, these species have a finite lifetime. The exciton, or the muon, can decay before it reaches its most stable state. So experiments may measure the properties of metastable states, and indeed may give very little information about equilibria. This natural metastability should be regarded as good news, in that the variety of excited states includes a wealth of different behaviours. There may be opportunities to manipulate and exploit the states and the possible branchings between competing excited state processes. The several consequences include solid state reactions, such as the photographic process, and enhanced diffusion in semiconductors.

Even in conventional systems, such as oxides or even diamond, equilibrium over charge states can be very slow indeed. X-irradiation of an insulating oxide, such as MgO, will cause transition metal ion dopants to change charge states. Non-equilibrium charge states may be observable—perhaps simply as colour changes—for months. The natural α- and β-irradiation of quartz in pottery causes ionic and electronic redistribution, and underlies the dating and authentication of art ceramics, some made several thousand years ago (Aitken 1985, Itoh *et al* 2001).

Massive excitation.   Massive excitation can be achieved that can even suffice to destroy the initial solid (Vaisburd and Balychev 1972). Such cases are rare, but it is a convenient metaphor in radiation damage to say that the centre of a collision cascade becomes liquid, since the energy deposited can be tens of electronvolts per atom. How good is the 'liquid' description? How well does it describe the more complex, but important, cases such as Si, where solid Si is a semiconductor and liquid Si is a metal? Can one pretend that the system is neutral everywhere, even though electrons will have been scattered out of the central 'liquid' zone to make it electrically positive? If there is such a transiently positive core region, how soon is neutrality restored? And how quickly and where will the electrons deposit their energy? The Fermi level is not defined, so it is hard to know how to decide what to do about charge states. For some of these features, a mesoscopic (or even continuum) description is effective. These mesoscale phenomena cannot be ignored in atomistic treatments of the central region.

These mesoscale issues become important in understanding the formation of permanent tracks (surviving for geological times) following the passage of swift heavy ions. Depending on the system, a range of ideas apply (Itoh *et al* 2009). For systems where thresholds are relatively low (oxides, some halides, amorphous semiconductors), excitonic mechanisms seem appropriate, with self-trapping and energy localization playing roles. For other systems (metals, crystalline semiconductors, oxides such as MgO and Al_2O_3) self-trapping is not an option, and the issue is whether energy can be retained along the track path long enough for a new structure to emerge. There are thus strong correlations with electronic thermal conduction and specific heat in metals (Duffy *et al* 2008). Note that defect annealing can occur, an example of a history-dependent process (figure 2).

Timescales.   Condensed-matter processes occur on many different timescales. Among the fastest are electronic relaxation in metals (~femtoseconds); photochemistry and fast
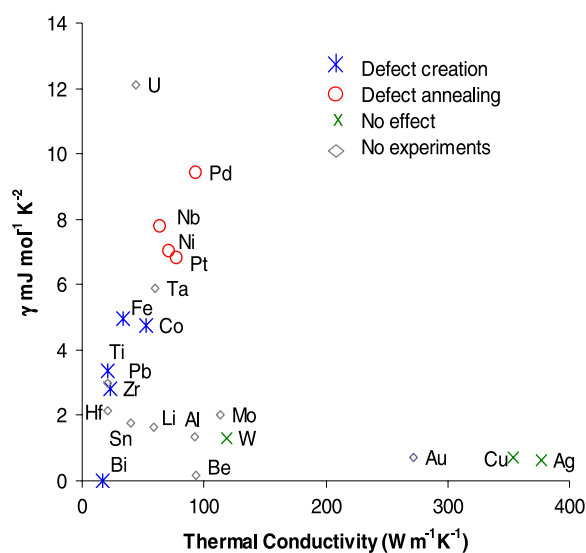
Figure 2. Electronic phenomena following massive electronic excitation (reprinted from Duffy *et al* (2008)) Excited electrons store energy that is only slowly transformed into atomic kinetic energy: the extent to which the electrons can do this is measured by the electronic specific heat. Electrons also move energy in space: a measure of this is the electronic or thermal conductivity.

non-radiative processes (∼0.1–1 ps). Slower, yet still rapid, are allowed radiative transitions (say ∼10 ns). Much slower processes include spin-forbidden transitions (∼milliseconds) and the diffusion-controlled processes that can extend over geological times. For steels in a nuclear reactor, the fastest processes in collision cascades take only a few femtoseconds, and most of the basic actions are over in a few picoseconds. But the consequences of these cascades and the timescales for subsequent diffusion processes to affect the ductility of the steel can grow in importance over 30 years, the billion second timescale. The development or not of ductility may affect reactor economics profoundly.

Life processes and metastability. Life processes, as Schrödinger (1951) noted, are inherently non-equilibrium. Many processes in living organisms are closer to a dynamic steady state than to equilibrium, and the hazards of living, plus the two major signatures of life (metabolism and reproduction) ensure that any equilibria are local. There are many striking phenomena that are hard to mimic, let alone model. For example, is Davydov's model (Davydov 1973, 1974, 1977) of efficient energy transfer along the α-helix of a protein by a soliton mechanism correct (Austin *et al* 2003, Lindgard and Stoneham 2003)? Modelling an α helix in isolation is not easy, though the simpler calculations, such as structure, and possibly hydration, appear solvable if computer power rises. Whether the coherent soliton motion—relying as it does on the consequences of many interactions between subunits—is as straightforward is hard to know. How energy is injected into the α helix is also not obvious, and raises one of the tantalizing questions at the physics/biology boundary. How are we to relate the biologist's views of force (typically as an energy divided by a distance or as an analogue of a Brownian force) to the physicist's clear ideas on forces as derivatives of potentials, acting at well-defined sites (Fisher and Kolmeiski 1999)? True, there is some common ground in irreversible thermodynamics. It must be added as well that entropic terms, rather than enthalpic, can dominate in biological situations. Modelling biosystems, going beyond mimicry to understanding, is an open field. There can be significant process by asking questions like 'is there a physics-based objection'

to some hypothetical mechanism. A useful example here is the question of how receptors are actuated very selectively by small molecules (e.g. Brookes *et al* 2007). Standard shape-based ideas ('lock and key') fail; the more general swipe-card descriptions (a swipe card or key card, such as a credit card, has to have the right shape to fit the reader, but the information transferred is not shape-based) need careful physics analysis. These phenomena sound very far from oxides and semiconductors. Yet the mineral components of life forms from diatoms to vertebrates are rich in oxygen, and their development into strong and functional skeletal components is controlled with precision and selectivity by soft matter.

4. Non-adiabatic and non-radiative transitions

Non-radiative transitions are remarkably varied. Some—such as thermal diffusion—take place near equilibrium, with only a single adiabatic energy surface relevant. Others specifically relate to highly non-equilibrium situations after excitation, and may involve several distinct adiabatic energy surfaces. Further examples concern loss of quantum coherence or entanglement, and are more subtle. This section identifies a few of the interesting (sometimes worrying) cases. Where do we need non-radiative transitions?

Irreversibility.   Even for the commonest non-radiative transitions, the non-adiabatic processes by which many excited systems revert to the ground state, quantitative predictions present significant difficulties (Itoh and Stoneham 2001). The normal input to such transition rates involved two components. One component is an electronic matrix element of the non-adiabatic part of the Hamiltonian, and this needs wavefunctions rather than charge densities. Difficulties have been identified, some subtle, but there are few attempts to do state-of-the-art calculations. One potential difficulty is this. When we consider optical excitation, there is a clear distinction between that part of the Hamiltonian driving the transition and that part representing the system that is changed. When there is a non-radiative transition, the partitioning of the Hamiltonian can be ambiguous, though most difficulties disappear when one ensures that the theory matches experiment properly. The second component of the transition rate concerns the quantum treatment of nuclear motion, going beyond the usual classical dynamics with quantum treatments of electrons. Unfortunately, the usual Fermi Golden rule, or something broadly equivalent, does not suffice to predict the evolution of a system by self-consistent molecular dynamics. This is not a problem of the harmonic approximation, or of unusual adiabatic dynamics, such as solitons in conducting polymers, all of which can be handled. The non-adiabatic aspects are tricky, elusive in a convenient form. What is needed here is a means to avoid the Fermi Golden rule. Various tools, such as frozen Gaussian methods or energy surface hopping, offer ways forward, but there is little theory that addresses the remarkable experiments on transient excited state phenomena in halides and oxides, for example.

Multiple energy surfaces.   Situations involving multiple energy surfaces are challenging. It is only too tempting to treat a Jahn–Teller system as if there were a single energy surface. In some cases, this crude approximation may be acceptable. Yet this would be wrong dynamically, and would also omit extra degrees of freedom and their potential excitations, such as orbitons in coupled Jahn–Teller systems (e.g. Sarfatt and Stoneham (1967)). Such coupled systems can exhibit other ideas hard to simulate, such as frustration, or distinctions between real and quasi spins when the interactions between component magnetic moments are significant. The complex oxides, often perovskites, show a wealth of behaviour mixing Jahn–Teller, polaronic and many-electron phenomena.

Quantum electrons and quantum nuclei.   The major success of theory in recent years has been self-consistent molecular dynamics, pioneered by Car and Parrinello (1985). Their approach assumes quantum electrons and classical nuclear motion. The quantum description of the nuclear motion emerges in several cases, some already mentioned, and most frequently in the difficult area of quantitative descriptions of tunnelling.

Complexity.   It goes without saying that some of the most significant areas to address involve a complicated mix of behaviours. Complications alone do not necessarily imply complexity, in the formal sense, in that there may be simple effective solutions where choosing the appropriate first approximation may lead to success. But other situations are far from rare. These include situations of special sensitivity (e.g. near phase transitions), cases where rare events dominate (e.g. electrical breakdown or the emergence of a superconducting path in a ceramic superconductor) or non-linear systems, including catalytic behaviour.

5. Full quantum treatments, including predictive decoherence

Serious predictions of electronic structure inevitably involve quantum electrons, the specifically quantum behaviour arising through the Planck constant $\hbar$ in kinetic energy or momentum operators, through quantum statistics with $\hbar\omega/kT$ factors, or through the exclusion principle. When the adiabatic approximation is not enough, quantum nuclear motion is needed, as well as that of electrons. In some highly correlated systems, and in superconductivity, the quantum effects become both more sophisticated and less intuitive. Indeed, it is not clear what practical limits exist to the modelling of quantum systems. The ways quantum features are evident are varied. Sometimes it is simply discreteness of energies, e.g. an electron may only couple to vibrations of one frequency, and can thus lose energy only in specific units. Sometimes it is coherence of a state, e.g. in excited states of colour centres (Itoh and Stoneham 2001) or in the excitation of photosynthesis centres (Engel *et al* 2007), the system may make several coherent vibrations between each emission of a vibrational quantum. But the quantum feature may be the more subtle entanglement.

Quantum information processing, and the even more demanding ideas of quantum computing, are widely discussed (e.g. Williams and Clearwater (2000)). In particular, the power of quantum computing and some quantum cryptographies lies in entanglement, the quantum dance of one electron with another and the way that is reduced by decoherence, the quantum analogue of classical dissipation. Entanglement goes beyond simple coherence. Interestingly, these aspects of quantum behaviour may require the quantum system to stay away from equilibrium. Quantum statistics, at its most useful at or close to equilibrium, has $\hbar\omega/kT$ factors that suggest that the consequences of the quantum $\hbar$ can be defeated by high temperature T. If one is to have a room temperature quantum computer, one must stay away from equilibrium: quantum entanglement, and the manipulations of quantum information processing, needs highly non-equilibrium situations. And, at some stage, it will be important to predict entanglement. For model systems, this is possible. Understanding the extent of the problems is a first step to predictions in important cases.

6. Open systems and interfaces between unlike materials

Many important interfaces involve dissimilar materials: a metal and an oxide; a polymer and an indium tin oxide electrode; an adhesive and wood; a biomaterial and blood. Some of the challenges are to understand mechanical properties, such as adhesion and friction. Sometimes

simple ideas work well, such as the image interaction picture of non-reactive metal/oxide adhesion (Stoneham and Tasker 1985), and can give at least a framework for discussion of complex interfaces, such as that between spent oxide nuclear fuel with alloy clad. The role of topography should not be ignored, since it is this that enables striking natural phenomena, like the way lotus leaves repel water or the way a gecko moves over a ceiling. The key ideas can be counter-intuitive. For instance, when using a polymeric adhesive, the work of adhesion is usually determined by the many weak bonds as polymer strands are pulled through rather than the strong surface/polymer binding (de Gennes 1990). And, if a chemical reaction occurs between an oxide and a metal so that an intermediate compound is formed, one needs to decide whether it is the interface between the oxide and the intermediate, or the metal and the intermediate, that determines failure.

Other issues concern the processes of transferring charge (electrons, protons, etc) across the interface. It is not simple to match absolute energies on the two sides of most interfaces. There are often significant dipole moments, for instance. Charge transfer, whether through a blocking electrode or through an ohmic contact, often implies an *open* system, whether it be a biosystem and its environment, or a nanocomponent within a device. In such cases, the boundary conditions demand subtle treatment. Moreover, the dynamics of screening can take one well away from the usual classical electrostatic descriptions. Tunnelling rates and injection phenomena need a dynamic description of screening (Ness and Fisher 1999, 2002, Ness *et al* 2001, Stoneham *et al* 2002). Electron emission needs to recognize the long-range electric field, and theory is largely phenomenological, a situation that is becoming increasingly unsatisfactory as nanoscale emitters are evolving.

Other open systems are those in which growth occurs. Thus the oxidation (or its prevention, as in passivation) of metals or semiconductors includes early stages of nucleation and what one must consider as surface chemistry, and then a series of growth mechanisms that depend on competing processes and possibly on electrostatics. In biological systems, the way that ordered structures with functionally effective structures emerge is still understood only weakly (e.g. Stoneham (2007)).

7. Mesoscopic issues where rare events dominate

The gate dielectric has been a critical component of silicon semiconductor devices. In the early 1950s, silicon was not the semiconductor of choice, but it soon became clear that its oxide gave it remarkable advantages. Silicon dioxide, the usual gate dielectric, has three main virtues. First, it passivates surface states on silicon. Secondly, it forms an amorphous but essentially stoichiometric oxide that is a good dielectric (an ideal dielectric needs other properties, of course, which is why the International Semiconductor Roadmap indicates that silica is giving way to hafnia; see, e.g. Houssa (2004)). Thirdly, it can be used in silicon processing. As a dielectric, one key property is breakdown. But how does this happen? The vast range of experimental studies (Lombardo *et al* 2005) suggests that defects develop in the electrically stressed oxide, that these ultimately form a percolation path through which a current can flow and irreversible damage develops further. The description just given is pathetically weak in detail. In particular, the statistics of breakdown, showing extremal rather than normal character, suggest that damage is initiated at rare structural features. This, of course, is also the case for fracture or for electrical breakdown in other standard insulators, such as polymers or other oxides. One measure of the rarity is this: at breakdown, the polarization energy stored per atom is of order 1/10 000 of the energy to break one bond (interestingly, the elastic energy stored just before fracture is of the same order). Suppose one wants to model electrical or mechanical breakdown. Standard approaches do not suffice, for fracture is qualitatively

distinct from modest shifts in elastic constants. How does one create a model system that will give the probability of these rare events properly? The reversible switching of oxides shares some features with oxide electrical breakdown. Polymer electrical breakdown has new complexities, partly from disorder, and partly because many insulating polymers have negative electron affinity: excited electrons have lower energies *between* strands than on a polymer strand. The theory of negative electron affinity systems is in its infancy, yet is of considerable economic importance.

In some condensed-matter processes, such as collision cascades or fast heavy ion bombardment, energy densities can be extremely large, perhaps tens of electron volts per atom. In such cases, and even under irradiation at lower energies, there will be a hierarchy of processes. In polymers, the irradiation will cause scission or cross-linking, and these changes will influence subsequent damage mechanisms. Thus low energy nitrogen ions can knock out hydrogen atoms from a hydrocarbon, the repeated cross-linking leading to a diamond-like carbon, a process that can be modelled (Godwin *et al* 1996, Stoneham *et al* 1998). Or interstitial molecular oxygen may determine whether a broken bond simply recombines or is prevented from rebonding by reaction with the oxygen. Key features of radiation behaviour are therefore described by radiochemical yields (what happens for every 100 eV put in?), that have to include secondary effects. These are matters of importance, since a polymer's embrittlement and tendency to fracture, plus a reduced electrical breakdown performance, is critical for applications such as insulation in fission reactors. Yet fracture and breakdown predictions are classic examples of behaviour for which phenomenological behaviour is well documented, but reliable predictions, even of behaviour statistics, rarely attempted.

When there are hierarchical processes, one common symptom is non-Debye relaxation behaviour (stretched exponential, 'universal response', etc). Such behaviour is probably a consequence of the extremal statistics already mentioned, though possibly involving several statistically distributed phenomena. And, if one is dealing with fracture, or electrical breakdown, for which extremal statistics apply, how do you ensure your model has the right Weibull parameters? Focussing one's model on the *appropriate* scale seems much more effective (Stoneham and Harding 2003). Whilst multiscale modelling is not uncommon, there should be real doubts about the wisdom of using an all-embracing code that incorporates sophisticated calculations at several levels, e.g. finite difference macroscopic, atomistic and electronic structure. One problem is making sure the ideas are right at each stage: in a silicon MEMS device, for instance, is the atomistic treatment of the silicon to have priority over treatment of thermal oxide?

8. Posing the problem: making contact with reality

Theoretical materials science and technology has several levels, and also several roles. It provides a framework in which to organize empirical results. It can be used to scope a new field. It can be used to separate out the components of some complex system, where experiment alone still confuses. And one can imagine cases—especially for the shortest or the longest timescales—where theory can outreach experiment.

For all the early days of materials modelling, there was an implicit assumption that either the computer hardware or the software was the most serious limitation. Posing the problem effectively, and interpreting the results properly, were supposed straightforward. This no longer seems to be the case. The emphasis has changed. Major developments in hardware and software are welcome, but the serious limits need brainware and experience, rather than computer power or software (see, e.g. Stoneham *et al* (2001)). Can we frame the problem in a way that could be modelled at the *right* level of detail, precision and sophistication? Do we

understand enough of the modelling output to give adequate answers to key questions? Have we the underlying science to recognize the limitations of the models used and be aware of the value of the answer assessed? Do we have enough confidence to use our models to take unpopular decisions? Could we tell an influential person that their favourite idea can't work? Anyone who has worked in a technology-based organization will know that new and unexpected situations arise with remarkable regularity. Sometimes these are problems, sometimes opportunities. Usually, the window of opportunity is short, since timescales are fixed by non-scientific constraints. Posing the problem in a soluble form can be the biggest challenge. It can also be rewarding. As thermodynamicist J Willard Gibbs acknowledged, thermodynamics owed more to the steam engine than the steam engine owed to thermodynamics.

Acknowledgments

References

Aitken M J 1985 *Thermoluminescence Dating* (London: Academic)
Austin R H, Xie A, van der Meer L, Shinn M and Neil G 2003 *J. Phys.: Condens. Matter* **15** 1693
Brookes J C, Hartovtsiou F, Horsfield A P and Stoneham A M 2007 *Phys. Rev. Lett.* **98** 038101
Camara C G, Escobar J V, Hird J R and Putterman S J 2008 *Nature* **455** 1089
Car R and Parrinello M 1985 *Phys. Rev. Lett.* **55** 2471
Davydov A S 1973 *J. Theor. Biol.* **38** 559–69
Davydov A S 1974 *Biophysics* **19** 684–91
Davydov A S 1977 *J. Theoret. Biol.* **66** 379–87
de Gennes P 1990 *Liquides aux Interfaces* ed J Charvolin *et al* (Amsterdam: North-Holland) p 296
Dickinson J T, Jensen L C, Webb R L, Dawes M L and Langford S C 1993 *J. Appl. Phys.* **74** 3758
Duffy D M, Itoh N, Rutherford A M and Stoneham A M 2008 Making tracks in metals *J. Phys.: Condens. Matter* **20** 082201
Engel G S, Calhoun T R, Read E L, Ahn T K, Mancal T, Cheng Y-C, Blankenship R E and Fleming G R 2007 *Nature* **446** 782–6
Fisher M E and Kolmeiski A B 1999 *Proc. Natl Acad. Sci.* **96** 6597–602
Godwin P D, Horsfield A P, Stoneham A M, Bull S J, Ford I J, Harker A H, Pettifor D G and Sutton A P 1996 *Phys. Rev. B* **54** 15785–94, 348
Houssa M (ed) 2004 *High-k Gate Dielectrics* (Bristol: Institute of Physics Publishing)
Itoh N and Stoneham A M 2001 *Materials Modification by Electronic Excitation* (Cambridge: Cambridge University Press)
Itoh N, Stoneham D and Stoneham A M 2001 *J. Phys.: Condens. Matter* **13** 2201–10
Itoh N, Duffy D M, Khakshouri S and Stoneham A M 2009 *J. Phys.: Condens. Matter* accepted
Johannissen L O, Scrutton N S and Sutcliffe M J 2008 *J. R. Soc.: Interface* **5** S225
Lindgard P-A and Stoneham A M 2003 *J. Phys.: Condens. Matter* **15** V5–V9
Lombardo S, Stathis J H, Linder B P, Pey K L, Palumbo F and Tung C H 2005 *J. Appl. Phys.* **98** 121301
Löwdin P-O 1963 *Rev. Mod. Phys.* **35** 724
Ness H and Fisher A J 1999 *Phys. Rev. Lett.* **83** 452
Ness H and Fisher A J 2002 *Europhys. Lett.* **57** 885
Ness H, Shevlin S A and Fisher A J 2001 *Phys. Rev. B* **63** 125422
Sarfatt J and Stoneham A M 1967 *Proc. Phys. Soc.* **91** 214
Schrödinger 1951 *What is Life?* (Cambridge: Cambridge University Press)
Shaw M J, Briddon P R, Goss J P, Rayson M J, Kerridge A, Harker A H and Stoneham A M 2005 *Phys. Rev. Lett.* **95** 105502 (note the subsequent erratum)
Stoneham A M 1975 *Theory of Defects in Solids* (Oxford: Oxford University Press)
Stoneham A M 1989 *Nucl. Instrum. Methods Phys. Res. B* **48** 389–98 (Aarhus 265)
Stoneham A M 1996 *Radiat. Eff.* **142** 191–203
Stoneham A M 2005 Plenary Perspective: Where are the gaps? *Handbook of Materials Modelling* vol 1 (Methods and models) ed S Yip (Berlin: Springer) pp 2731–6

Stoneham A M 2007 *Rep. Prog. Phys.* **70** 1055–97

Stoneham A M, English C A and Phythian W J 1996 *Radiat. Eff. Defects Solids* **144** 311–34

Stoneham A M, Godwin P D, Sutton A P and Bull S J 1998 *Appl. Phys. Lett.* **72** 3142

Stoneham A M, Marta M D Ramos and Ribeiro R M 1999 *Appl. Phys.* A **69** S81–6

Stoneham A M, Howe A and Chart T 2001 Predictive materials modelling *UK Department of Trade and Industry/Office of Science and Technology Foresight Report* DTI/Pub5344/02/01/NP, URN 01/630 (A follow-up DTI Materials Innovation and Growth Team document is *Materials Modelling—Growth and Innovation Prospects* Stoneham A M and members of the EPSRC Materials Modelling Strategic Task Group 2006)

Stoneham A M, Gavartin J, Muñoz Ramo D and Shluger A 2006 *Phys. Status Solidi* a **204** 653

Stoneham A M, Ramos M M D, Almeida A M, Correia H M G, Ribeiro R M, Ness H and Fisher A J 2002 *J. Phys.: Condens. Matter* **14** 9877–98

Stoneham A M and Harding J H 2003 *Nature Mater.* **2** 77–83

Stoneham A M and Tasker P W 1985 *J. Phys. C: Solid State Phys.* **18** L543

Thompson D'Arcy W 1917 *On Growth and Form* (Cambridge: Cambridge University Press)

van der Kamp M W, Shaw K E, Woods C J and Mulholland A J 2008 *J. R. Soc.: Interface* **5** S173

Vaisburd D I and Balychev I N 1972 *Sov. Phys.—JETP Lett.* **15** 380

Webb R L, Jensen L C, Langford S C and Dickinson J T 1993a *J. Appl. Phys.* **74** 2323

Webb R L, Jensen L C, Langford S C and Dickinson J T 1993b *J. Appl. Phys.* **74** 2338

Watkins G D, Corbett J W and McDonald R S 1982 *J. Appl. Phys.* **53** 7097

Williams C P and Clearwater S H 2000 *Ultimate Zero and One* (New York: Copernicus)