Hybrid atomistic simulation methods for materials systems

Rep. Prog. Phys. **72** (2009) 026501 (25pp) doi:10.1088/0034-4885/72/2/026501

Hybrid atomistic simulation methods for materials systems

N Bernstein[1], **J R Kermode**[2,3] **and G Csányi**[3]

[1] Center for Computational Materials Science, Naval Research Laboratory, Washington, DC 203755343, USA
[2] TCM Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge, CB3 0HE, UK
[3] Engineering Laboratory, University of Cambridge, CB2 1PZ, UK

E-mail: noam.bernstein@nrl.navy.mil, jrk33@cam.ac.uk and gc121@cam.ac.uk

Received 3 September 2008, in final form 2 December 2008
Published 16 January 2009
Online at stacks.iop.org/RoPP/72/026501

Abstract
We review recent progress in the methodology of hybrid quantum/classical (QM/MM) atomistic simulations for solid-state systems, from the earliest reports in 1993 up to the latest results. A unified terminology is defined into which the various and disparate schemes fit, based on whether the information from the QM and MM calculations is combined at the level of energies or forces. We discuss the pertinent issues for achieving 'seamless' coupling, the advantages and disadvantages of the proposed schemes and summarize the applications and scientific results that have been obtained to date.

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

This article was invited by Professor M Finnis.

Contents

1. Introduction

In trying to simulate the behaviour of materials on the atomic scale, the first and most obvious compromise one has to make is to restrict the size of the simulated system. On supercomputers easily accessible to the academic community today, it is possible to carry out first principles simulation of hundreds of atoms for tens of picoseconds. Dropping the accuracy to the level of empirical interatomic potentials allows the simulation of millions of atoms for microseconds. This still falls far short of the length scale and especially the time scale of experiments. The usual strategy is to sidestep this shortcoming by only considering properties of the system that are converged with respect to the system size at the sort of system sizes that are available at the chosen level of accuracy. For some properties, convergence is very fast, e.g. many properties of point defects in periodic crystals can be accurately calculated by considering just a few hundred atoms [1].

There is a special class of properties (or rather, a special class of systems with many such properties), where this convergence is especially slow. These so-called strongly coupled multiscale systems are typically associated with a

long-range interaction, either electrostatic or elastic. The long-range nature of these forces means that the behaviour of a small region where some process is taking place can be affected by the collective configuration of a very large number of atoms over long, sometimes even macroscopic distances. Furthermore, for some systems, the converse is also true: the detailed interactions of certain key atoms can influence the equilibrium configuration of potentially tens of thousands of atoms.

For a system where elastic interactions dominate, the long-range interactions are indirect: if a bit of matter moves, only adjacent bits feel the effect immediately. In an atomistic description this is equivalent to real-space localization of the force-constant matrix (the derivatives of atomic forces with respect to atomic positions). However, if the system is allowed to relax, even material very far from the perturbation can be affected. The effect is particularly long ranged in the presence of a formal divergence of the stress field, for example, near a topological defect such as a dislocation or a sharp crack. With such a divergence a local rearrangement can lead to the relaxation of a macroscopic amount of energy throughout the system. The accurate description of the atoms near the divergence needs, therefore, to be augmented by an explicit inclusion of a large part of the system farther away. However, since the interaction is indirect, only the constitutive properties of the relevant field (i.e. elastic constants for continuum elasticity) of the rest of the system affect its interaction with the near-field region. Even if other formal singularities exist in the augmentation part of the system (e.g. a forest of background dislocations), the fact that an approximate atomistic (or even continuum) description is used there does not affect their long-range fields and therefore their interaction with the accurately described atomistic region.

For systems where Coulomb interactions are significant, both direct and indirect long-range interactions can occur. Direct long-range interactions arise from the $1/r$ distance dependence of the Coulomb interactions between point charges. If a charged atom moves, atoms far away on the atomic scale directly feel the effect. Screening may reduce the range of this interaction, but it nearly always extends to distances that would be hard to describe in a fully quantum mechanical simulation. In addition, the screening of the local charge or charge multipole can involve the collective motion of charges in a large (on the atomic scale) neighbourhood around it, leading to an indirect interaction analogous to the relaxation of atomic positions in a non-polar material subject to a local perturbation. As for continuum elasticity, the accurate description of the material in a small region requires that a large region around it be described explicitly in the simulation. For the electrostatic case the continuum property that controls the indirect interaction is the polarizability, and an analogy can be made to the elastic constants for continuum elasticity. However, since there are also direct interactions, approximations that preserve the polarizability (analogous to potentials that give the correct elastic constants for embedding a system where elastic interactions dominate), such as shell-model potentials, are not necessarily sufficient. Here the interacting particles themselves are associated with a formally divergent field, and so making an error in the charge or the dipole moment of an atom in the augmentation region will give rise to errors in its interaction with the accurately described region. Interatomic potentials do typically make these errors, in contradistinction to the elasticity case, where the topological fields that control the indirect coupling are described exactly.

A somewhat different scenario for using hybrid modelling exists in the case of systems where long-range coupling is one way only, namely, the collective far-field atomic configurations affect the chemistry in a specific region, but the atomic fluctuations due to a finite temperature play a significant role. In particular, this is the reason for the use of hybrid techniques in the fields of physical chemistry and biochemistry, where one is interested in the details of chemical reactions of solutes in polar solvents or of ligands at the active sites of proteins. But the same issues can arise in connection with impurities and defects in and on the surface of ionic materials. In these cases, the correct distribution of the long-range fields at the chemically changing site could, in principle, be computed in advance and used as stochastic boundary conditions for a chemically accurate small scale simulation. We are not aware of published works along these lines, and instead all studies to date resort to representing the far-field atoms explicitly.

The need to compute material properties at finite temperatures is also the reason why the far-field atoms need to be represented at all in the case of elastic coupling. For zero-temperature properties, it is natural, as is being done by several groups, to couple the quantum mechanically described region directly to a continuum model, e.g. finite elements [2–4]. However, to date, there has not emerged a satisfactory treatment of heat exchange (i.e. vibrations) between regions described atomistically and by finite elements. The root of the problem is that in the finite-element region, in areas where the element size is close to the atomic length scale, some of the heat is represented by the explicit kinetic energy of the nodes and some by the heat field variables internal to each element. The two representations account for the energy stored in phonons with wavelengths that are larger and smaller than the element size, respectively. We are not aware of any attempt to treat the heat flow between these two representations as the element size varies on the mesh.

In this review, we focus exclusively on hybrid schemes that combine two atomistic treatments: one explicitly quantum mechanical and one using interatomic potentials (or the so-called 'force field', often labelled 'classical', although it must be pointed out that any 'classical' interatomic model is still an approximation to the underlying quantum mechanics, albeit perhaps a crude one). We adopt the nomenclature of the chemistry literature and refer to the former as QM and the latter as MM (which stands for 'molecular mechanics'). The distinguishing feature of the MM models is that no electronic degrees of freedom representing metallic or covalent bonding are present. Almost universally, the total energy of the system is written as a sum of atomic site or interatomic bond energies, which are analytical functions of the coordinates and charges of the other atoms. An exception to this is the handful of variable charge and polarizable models, which involve the computation of the charge or dipole moment of each atomic site using a self-consistent procedure [5, 6].

The goal of a hybrid simulation is to give the same observables as one would measure in a fully quantum mechanical simulation. Therefore this must form the basis for evaluating various hybrid schemes. Equilibrium geometric properties are generally easy to obtain to such accuracy; a much more stringent test is whether energy and free energy differences are reproduced.

2. QM region identification

A natural question that comes up when dealing with QM/MM hybrid schemes is which parts of the system should be treated with quantum mechanics. There are several closely related issues in this regard. Firstly, general decisions need to be made about which type of event or atomic configuration should trigger a more accurate treatment and how many atoms should be treated quantum mechanically. Following this, the atomic events need to be tracked as they unfold.

The trade-off is simple: the more atoms are treated quantum mechanically at more numerous sites, the slower the simulation will become due to the typically overwhelming dominance of the quantum mechanical calculation in resource requirements. However, if not enough atoms are treated accurately, there is a risk that important events will be missed or that significant quantum mechanical interactions will be neglected.

Indeed, one naive approach would be to attempt a QM treatment for all atoms for which the MM method is deemed to fail. If the difference between QM and MM forces on the atoms is used to define 'failure', it turns out that widely available interatomic potentials never reproduce the quantum mechanical forces with any defensible accuracy at useful finite temperatures (at or above room temperature). Rather, one usually resorts to structural and geometric criteria: non-reactive interatomic potentials will certainly fail close to bond forming and bond breaking events, so these need to be captured using the QM model. The fact that the interatomic potentials also depart from the QM potential energy surface for routinely encountered fluctuations that do not lead to bond rearrangement is less important, although it certainly affects properties such as thermal expansion and thermal conductivity that depend on anharmonic parts of the potential.

It is thus clear that there cannot be one universally correct method for selecting QM regions, but that one must adapt heuristics for the particular observables that are being calculated. In almost all applications to date, the trigger for using quantum mechanics was self-evident: a topological defect such as a crack tip or a dislocation core or a foreign atom is introduced into the system and it and its immediate surroundings are treated by the QM model. The physical justification for this choice is that the stress field, which drives the evolution of these systems, is strongly concentrated at such defects and therefore non-trivial bond rearrangement will happen there first. In the following, we will denote atoms which trigger QM treatment as *active atoms*. Since the QM description leads to interactions of finite but limited range (except for classical long-range electrostatic effects), to correctly simulate their behaviour some neighbours of the active atoms must also be treated with a QM model. We define the union of these sets as the *QM atoms* which occupy the *QM region*.

Given some active atoms, the first question is how many nearby atoms should be treated quantum mechanically? The answer follows from the underlying goal of hybrid schemes, namely, that macroscopically the system should behave as if it was modelled fully quantum mechanically. The number of QM atoms is a control parameter, and macroscopic observables of interest should be converged to the desired accuracy with respect to this parameter. In practice, the actual control parameter would not be the number of QM atoms directly, but some auxiliary parameters of an algorithm which determines the number of QM atoms (e.g. the radius of a sphere centred on an active atom within which all the atoms are selected for QM treatment). Using such criteria enables the cheapest possible simulation while still giving correct answers.

When designing the algorithm that selects the QM region, it is important to minimize fluctuations in the set of active atoms and also in the complete set of QM atoms (for reasons detailed below). One trivial way of doing this, which is the most widespread, is to fix the set of QM atoms at the beginning of the simulation. While this might work for investigating a chemical reaction between a few specific atoms, it is not suitable in cases when the QM region needs to be moved during the simulation (e.g. because it is the tip region of a propagating crack or the core of a moving dislocation) or updated because atoms move in and out of the vicinity of the active atoms due to diffusion.

Given that some geometric proximity criteria will be used to select QM atoms near the important region, an obvious way to reduce the fluctuations in the set of QM atoms is to perform the proximity analysis not on the instantaneous positions of the atoms but on time-averaged positions. For example, let

$$\tilde{\boldsymbol{x}}(t) = (1 - \exp(-\Delta t/\tau)) \sum_{n=0}^{\infty} \mathrm{e}^{-n\Delta t/\tau} \boldsymbol{x}(t - n\Delta t),$$

where Δt is the time step of the simulation, so that the position vectors $\tilde{\boldsymbol{x}}$ track the real positions $\boldsymbol{x}$ while filtering out fluctuations which have a timescale shorter than τ. Such filtering also makes the selection algorithm insensitive to small variations in the values of parameters such as nearest-neighbour cutoff distances. One limitation of this approach is that the time-averaged positions lag the current positions, so they will not instantaneously reflect changes in bonding.

With the low-pass filtering in place, one can further minimize fluctuations by using 'bond hopping' to define proximity, rather than Euclidean distances. In this case, the n-neighbourhood of an atom would be defined as 'all atoms that can be reached by hopping along nearest-neighbour bonds at most n times'.

Another important use for bond-hopping based selection is in organic materials, which can often be decomposed into compact covalently bonded groups interacting with each other via weaker forces, such as hydrogen or van der Waals bonding, and perhaps one or two inter-group covalent bonds. Using bond hops makes it possible to include or exclude entire groups,

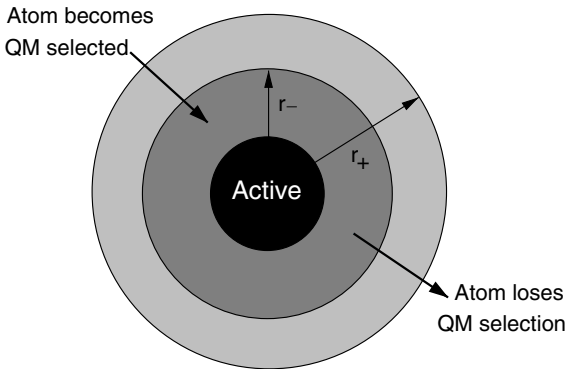

Figure 1. Hysteretic QM selection algorithm. The central (black) region contains the 'active' or 'core' atoms. Atoms must enter the region $r < r_-$ to become selected, and then remain selected until they leave the larger region, i.e. $r > r_+$. This has the effect of making the QM region insensitive to small fluctuations in atomic positions across either boundary.

minimizing the energy and force errors caused by breaking covalent bonds at the QM region edge.

An even more robust way to eliminate fluctuations is to employ 'hysteresis' in the selection process, as illustrated in figure 1. Roughly, this means that the criteria that atoms have to fulfil to become *selected* for QM treatment is more restrictive than the negative of the criteria to become *deselected*. For example, atoms have to come within a certain distance, r_-, of an active atom to become part of the QM set, but subsequently they have be further than r_+ to become MM again, with $r_+ > r_-$. With a sufficient gap between r_+ and r_-, the QM set can be made to change only when there is permanent topological change in the atomic configuration.

Similar ideas can be used not only to update the QM region around a given set of active atoms but also for updating the set of active atoms. For example, if it is decided that bond breaking and bond forming events should be tracked, then monitoring the number of neighbours of atoms, calculated based on the time-averaged positions $\tilde{\boldsymbol{x}}$, will detect permanent topological changes. All the above strategies combined were necessary to follow the moving tip of a crack with a QM region in [7].

3. Conservation laws

In the context of hybrid schemes, the question of Hamiltonian dynamics needs to be addressed. In conventional molecular dynamics (MD) the forces are derivatives of an energy function with respect to atomic positions, and therefore the total energy of the system should be conserved. Any deviations from this are signs of a mistake in the evaluation of analytical forces in the equations of motion or result from a time step that is too large for the particular integrator [8].

Before we discuss the energy conservation properties of various hybrid MD schemes, it is worth pointing out that conservation of energy is not always a desirable thing. In particular, when the QM region changes, there is a case to be made for not wanting to conserve energy. Suppose that we have a well-equilibrated system under a fully Hamiltonian scheme, and we suddenly turn on a QM region or the existing QM and MM status of atoms is updated. The atoms whose description is thereby changed from MM to QM or vice versa will now not be in equilibrium, and during their equilibration some net energy will be converted from kinetic to potential or vice versa. This will therefore heat or cool the system, which can change the dynamics in unexpected ways. This extra heat energy must be dissipated in order to return the system to the desired simulation temperature. Thus, while Hamiltonian dynamics confers its benefits to a hybrid simulation with unchanging QM regions, as soon as we wish to update the QM region, some kind of external thermostat is needed. Nevertheless, it is to be expected that a scheme that is energy conserving at least in the case of fixed QM regions is more stable from a practical point of view and requires weaker thermostats, which in turn are less likely to introduce artifacts.

Broadly speaking, atomistic hybrid schemes can be classified according to whether they combine the empirical and quantum mechanical models on the level of energies or forces. Basic schematic diagrams for each class are shown in figure 2. Those that build a combined total energy function can do so additively (by combining the QM energy, the MM energy and an interaction energy), subtractively (by combining the MM energy of the entire system with the difference between the QM and MM energies of the QM region) or by mixing local energies computed by QM and MM. All these approaches are naturally energy conserving, as long as forces are computed correctly. Hybrid approaches that are based on forces, on the other hand, combine QM forces for atoms in the QM region with MM forces in the MM region, possibly with an interpolation between the two in a transition region. These approaches may not conserve energy if the set of forces they end up with is not derivable from a potential energy function. Furthermore, if the QM and MM forces are directly used in the respective regions to drive the dynamics, as in the basic force-mixing scheme (see below), the dynamics will not conserve momentum. This violation of Newton's third law can be lessened by smoothly turning from using QM forces to MM forces across a very large transition region or entirely eliminated by simply shifting, for example, the QM forces by a constant such that the total momentum is conserved. Surprisingly, as we show below, this latter scheme results in much more accurate observables for a given QM region size (i.e. for a given computational cost).

4. Seamless coupling

As mentioned in the introduction, a coupled simulation should produce dynamics that yield the same observables as would be obtained if the whole system was treated with QM. Since the full QM system would not single out the region where the coupled simulation switches from QM to MM, coupling methods strive to eliminate any effects from the transition between the two models. This was referred to as 'seamless coupling' by some of the first developers of solid-state QM/MM coupling schemes [9]. Two conditions are necessary to achieve this seamlessness perfectly: matching of bulk QM and MM properties, and the elimination of edge effects in the separate QM and MM simulations that are combined into a complete coupled system. Here we use the term 'bulk' to

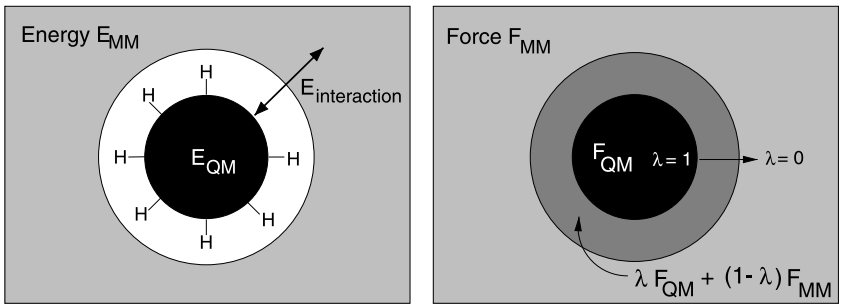

Figure 2. Schematic representations of the (*a*) energy-mixing and (*b*) force-mixing approaches.

loosely refer to the set of configurations for which the MM model will be used. Which bulk properties need to be matched will depend on the system and phenomenon of interest, for example, the elastic constants in a system dominated by elastic interactions. A simulation can, of course, be useful even if the coupling is not completely seamless. As long as the important properties are matched well enough, and as long as discontinuities at the QM/MM boundary are small enough, short ranged enough and far enough from where the interesting physics or chemistry is taking place in the QM region, the dynamics in the QM region may produce the correct physically observable properties.

4.1. Bulk property matching

The simplest way to express the requirement for perfectly seamless coupling is that the bulk dynamics produced by the QM and MM descriptions of bonding be identical. In the context of MD and Monte Carlo, this means that the energies be the same for the two methods, up to a uniform shift, for all bulk atomic configurations. Equivalently, the gradient of the energy with respect to atomic positions, i.e. the atomic forces, must be equal. Although clearly in practice such a requirement will never be met fully by the MM models, there have been some attempts to address it in the literature. The most basic and easiest to adjust property, the equilibrium lattice constant under the MM model for a perfect crystal, can easily be matched to that under the QM model by adjusting the length scale of the MM potential [10]. If the lattice constants do not match, a completely non-local artifact, namely, subjecting the QM region to a hydrostatic stress field, occurs. The other properties that are often mentioned as desirable to match are elastic constants [7, 11] and the Einstein oscillator frequencies, i.e. the energetics of the displacement of one atom with respect to the rigid lattice [9]. Both are attempts to capture the energetics of small displacements, in opposite length scale limits.

Since the role of the MM system is often to produce correct long-range elastic fields, matching elastic constants is a necessary condition: without it, even if the lattice constant is matched, as soon as the system is under uniform elastic deformation the QM region would respond with a different deformation. Matching the Einstein oscillator frequencies does not have as obvious a physical interpretation. It is however intuitively appealing as a proxy for the vibration dynamics of atoms on very short length scales. In general, both these approaches to matching dynamics are incomplete. Both focus on infinitesimal displacements, and each is restricted to one length scale extreme: infinite for elastic constants and the interatomic scale for Einstein oscillator frequencies. One possibility for going beyond these limitations is suggested by the proportionality between the elastic constants and the slope of the acoustic phonon dispersion near the vibrational Brillouin-zone at the Γ-point. The energy of displacements at all length scales, although still of infinitesimal amplitude, is captured by the full phonon spectrum or equivalently in real space by the force-constant (or dynamical) matrix. Matching the force-constant matrix, i.e. the first derivatives of the forces with respect to atomic positions, would guarantee a perfect match between the MM and QM forces for arbitrary infinitesimal displacements from equilibrium. While this would clearly be desirable in general, in the case of an energy-mixing scheme, only such a strict matching criterion can guarantee that spurious forces such as those observed in [12, 13] are eliminated exactly for near-equilibrium configurations. (The errors resulting from mismatching dynamical matrices are analogous to what are called 'ghost forces' [14] in the quasicontinuum [15] method, which is a well-known continuum/atomistic hybrid scheme.) Consider the force $\boldsymbol{F}_i$ on atom i, which can be written using the dynamical matrix, for a given energy model, as

$$\boldsymbol{F}_i = -\frac{\partial E}{\partial x_i} = -\sum_k \frac{\partial}{\partial x_k}\frac{\partial E}{\partial x_i}\Delta x_k$$

for small displacements Δx_k measured from equilibrium. Now the system is partitioned and two different energy models, E^{QM} and E^{MM}, are used to evaluate atomic energies. The energy of the total system is now

$$E = \sum_{j\in\mathrm{QM}} E_j^{\mathrm{QM}} + \sum_{j\in\mathrm{MM}} E_j^{\mathrm{MM}}.$$

Note that both E_i^{QM} and E_i^{MM} depend on positions of atoms in both the QM and MM regions, because in general the models are nonlocal. This corresponds to an abrupt change in the description of the system as we pass from one region to the

other. The same force is then given by

$$\boldsymbol{F}_i = -\sum_{j\in\mathrm{QM}}\sum_k \frac{\partial^2 E_j^{\mathrm{QM}}}{\partial x_k \partial x_i}\Delta x_k + -\sum_{j\in\mathrm{MM}}\sum_k \frac{\partial^2 E_j^{\mathrm{MM}}}{\partial x_k \partial x_i}\Delta x_k$$
$$= -\sum_k \left(\sum_{j\in\mathrm{QM}} \frac{\partial^2 E_j^{\mathrm{QM}}}{\partial x_k \partial x_i} + \sum_{j\in\mathrm{MM}} \frac{\partial^2 E_j^{\mathrm{MM}}}{\partial x_k \partial x_i}\right)\Delta x_k.$$

Note that, again due to nonlocality, there are non zero contributions from derivatives with respect to positions x_i and x_k of atoms in both regions. So the condition for the force to remain unchanged (so that it remains zero in the nominal equilibrium configuration) is

$$\frac{\partial^2 E}{\partial x_k \partial x_i} = \sum_{j\in\mathrm{QM}} \frac{\partial^2 E_j^{\mathrm{QM}}}{\partial x_k \partial x_i} + \sum_{j\in\mathrm{MM}} \frac{\partial^2 E_j^{\mathrm{MM}}}{\partial x_k \partial x_i}, \qquad (1)$$

which can be achieved in general only if the dynamical matrices of the two models are actually equal to each other and to that of the original single model. Even then, relation (1) is not guaranteed for atoms near the boundary, as it further depends on how the spatial partitioning of the energy is handled in the two models. In other words, the relation depends on how the *local* energies defined by the two models vary as a function of atomic displacements. Finally, even if we have the freedom to generate a new MM model, matching the dynamical matrix of the QM model is in itself a highly non-trivial requirement: the force-constant matrix is in general long ranged, and the most broadly applicable expressions have issues with rotational invariance or computational expense [16].

We have had some success in fitting an interatomic potential for silicon carbide with an intermediate approach [7]. Using only first-neighbour interactions (of Stillinger–Weber (SW) form [17], although this is not essential, since we are only fitting the location of the minimum and nearby curvature), we fit the lattice constant, elastic constants for long-wavelength information and the Raman mode, i.e. the Γ-point optical vibration frequency, for short wavelength information. For example, when targeting the density functional tight-binding (DFTB) QM model [18], we were able to match the lattice constant essentially perfectly, the elastic constants within 8% and the Raman mode frequency within 35%.

The matching of the MM and QM models for larger displacements is also important. Fluctuations to displacements beyond the quadratic regime increase with increased temperature. More importantly, thermal expansion is controlled by nonlinearities in vibrational properties, and at high temperatures a mismatch in thermal expansion between the QM and MM regions could lead to the same effects as a mismatched zero-temperature lattice constant. In principle, as long as the atomic configurations remain in the vicinity of the zero-temperature structure, higher order information can be easily used. Nonlinear elastic constants and nonlinear force constants (higher order derivatives of forces w.r.t. displacements) can be added to the fit, although the choice of functional forms becomes crucial.

Matching only elastic constants or any other property of infinitesimal displacements is thus useful, but limited. It seems sufficient superficially if one assumes that the interesting processes, which involve large displacements from the equilibrium state, occur only in the core of the QM region. However, such processes can easily leave behind structures that are locally stable, but far from the equilibrium state. If the QM region is to be kept compact, it needs to follow the action, and it may not be practical to keep all QM atoms in the QM region for the entire duration of the simulation. A crack, for example, may only require a QM description instantaneously where bonds are broken at the moving tip, but it leaves behind a free surface as it propagates. The surface reconstruction geometry in the QM region may have unphysically high energies or forces under the MM model. As the QM region moves to follow the crack tip, reconstructed surface atoms that switch back to the MM description might suddenly be subject to unphysical forces. No comprehensive solutions to these issues are known. An important conclusion is, however, that if interatomic potentials are generated specifically to be used in hybrid schemes, the criteria according to which their parameters are optimized should be quite different from those when making general purpose potentials.

4.2. QM region boundary effects

While matching bulk properties is strictly necessary, it is not sufficient for a seamlessly coupled simulation. Both the QM and MM calculations are usually carried out on only a subset of the atoms. Even though it is computationally feasible to simulate the entire system with the MM model, it is common to use a QM region that includes an atomic species that cannot be described by the MM at all and must therefore be excluded from the MM calculation. The QM calculation will obviously only include a small cluster surrounding the core QM region, since that is how the coupled simulation can be faster than the full QM treatment. The open boundaries between the QM and MM regions can affect the simulations, and these boundary effects must be minimized.

The way in which these boundaries can affect the computed MM and QM results (energies, forces) depends on the individual methods and on how the coupling scheme is formulated. The simplest case is for force mixing: as long as the MM (QM) forces on the atoms in the MM (QM) region are correct, there will be no boundary effects. This is in fact the major advantage of force mixing. For energy-based coupling, there are two possible sources of edge effects: double-counted or missing contributions to the total energy and incorrect energies contributed by passivation, link or buffer atoms that are used to lessen the effect of broken covalent bonds across the boundary. There are no systematic quantifications of these errors in the literature. We show a limited comparison in section 6.

4.2.1. Short-range interactions.

Force mixing. Getting the correct short-range contribution to forces for force-mixing coupling [10, 19, 20] is straightforward. The calculation (MM or QM) includes both the atoms where the forces are needed, plus a 'buffer' layer that gives

those atoms the correct environment. For short-ranged MM models, as long as the buffer layer is thick enough (typically twice as thick as the cutoff of the potential, for bond-angle dependent potentials), forces on the MM atoms are identical to those computed if every atom is included. In the QM region things are not as simple. The QM description of bonding is inherently non-local: the range of effects in the electronic structure is tied to the real-space decay of the density matrix, which is exponential in insulators and in metals at finite temperature [21–23]. This suggests that the adverse effects of the boundary of the region included in the QM calculation (QM atoms and buffer layers) on the QM atom forces could decay exponentially with the thickness of the buffer layer. In practice, however, several factors can complicate the behaviour of the system.

The strongest effect, particularly in insulators, is that the surface at the boundary of the QM region leads to dangling bonds and surface states. These states can lead to a closing of the band gap and to charge transfer, which can have relatively long-range effects. Passivating the dangling bonds using H atoms or specially tailored passivation pseudo-atoms, both commonly used in quantum-chemistry oriented QM/MM methods, can alleviate this problem to a substantial extent. When the passivation atom is designed to mimic the missing atom from the broken bond [9], it can be placed at the missing atom site. When the passivation atom is an H [11, 24], it is typically placed along the broken bond direction, but either at a fixed distance (e.g. the equilibrium X–H bond length) from the surface QM atom (X) or at a fixed fraction of the broken bond length. We and other practitioners of QM/MM coupling for solid-state systems have mainly used H for surface passivation [11, 20, 24, 25]. We have found that it works well for Si and C systems, for example [20]. Specialized passivation atoms optimized for a particular system have been shown to be very effective in organic chemistry applications [26]. However, the range of covalent bond chemistries present in solid-state systems of interest is much broader than in biological and organic systems, where one can nearly always choose to cut only bonds between a C and a C, O or N. Optimizing a passivation atom for a particular type of bond, for example, the Si–Si bond in the MAAD work [9], would not be useful for any other material system, and so optimized passivation atoms have not found widespread use in solid-state applications. Even fairly good surface passivation, such as that provided by H for Si and C systems, may be inadequate in some apparently simple systems. We have found that in minimal basis set methods that do not allow for much electronic screening, even small errors in matching the polarization of the broken bond can lead to arbitrarily long-range charge transfer and a complete failure of the forces in the QM region to converge as a function of the buffer layer thickness. In plane-wave density functional theory (DFT), which has the full variational freedom for the electrons, this artifact is much milder.

As an example, consider the silicon carbide system shown in figure 3. Using the semi-empirical DFTB method, we find that forces calculated on the central Si and C atoms oscillate as the cluster size is increased, converging very slowly to the result of a calculation with periodic boundary conditions (used as a benchmark, since it is free of boundary artifacts). Modifying the cluster surface by including an additional layer of carbon atoms has a pronounced effect on the long-range charge transfer, and the forces on the central atom no longer converge to the periodic result. Enforcing global charge neutrality in the QM region makes the two clusters consistent and results in much faster convergence, but does not converge to the periodic result either. Repeating these cluster calculations with a DFT code [27] shows no significant difference in the local charges and forces between the two cluster types.

Using a buffer layer and passivation atoms to give accurate forces in the QM region is a form of potential embedding, i.e. boundary conditions only in the electronic potential, as distinguished from 'quantum embedding', i.e. boundary conditions on the electronic states (wavefunctions or density matrix) themselves [28]. A version of quantum embedding is used in the DCET approach [10]. In this method the electronic density matrix is approximated as a sum of a small number of Green's functions (GFs), and the GF at complex energy z is computed by iterative matrix inversion of $z\boldsymbol{S} - \boldsymbol{H}$. The boundary conditions are applied by constraining the matrix inversion: any matrix elements involving tight-binding basis functions around atoms in a buffer region are constrained to the values they would have in the perfect crystal. This requires that the displacements in and near the buffer region be small enough to map particular atom pairs to equivalent pairs in the perfect crystal. This approach is in some ways more general than H atom passivation, since it should apply to any element and any type of bonding, although it is tied to the atomic orbital basis and to the GF-based solution of the electronic structure problem.

The optimal construction of the buffer layer has not received much attention in the literature. Since covalent bonds are primarily a first-neighbour interaction, most of the decay-range effects in covalently bonded systems are naturally interpreted in terms of first-neighbour hops. This suggests that the distance of a buffer layer atom from the QM atoms is better measured in terms of first-neighbour hops, rather than in real space. Another reason for preferring hops is that it significantly lowers the fluctuations in the set of buffer atoms, similarly to the set of QM atoms, as discussed above.

Energy-mixing. Energy-based coupling methods present a different set of challenges for minimizing the effects of the QM and MM region boundaries. In most such methods, the QM region is passivated but not buffered [9, 11, 24, 29, 30]. The one exception is Ogata's method [11, 24], which we discuss below. In the ONIOM approach [11, 24, 31] the total energy is defined as

$$E_{\mathrm{tot}} = E^{\mathrm{MM}}_{\mathrm{sys}} + E^{\mathrm{QM}}_{\mathrm{cluster}} - E^{\mathrm{MM}}_{\mathrm{cluster}}, \tag{2}$$

where $E^{\mathrm{MM}}_{\mathrm{sys}}$ is the energy of the entire system computed with the MM, $E^{\mathrm{QM}}_{\mathrm{cluster}}$ is the energy of the passivated cluster computed with QM and $E^{\mathrm{MM}}_{\mathrm{cluster}}$ the energy of the passivated cluster computed with MM. This method is illustrated in figure 4. Because the forces are derivatives of the energy, a similar equation holds for the forces, although with forces on

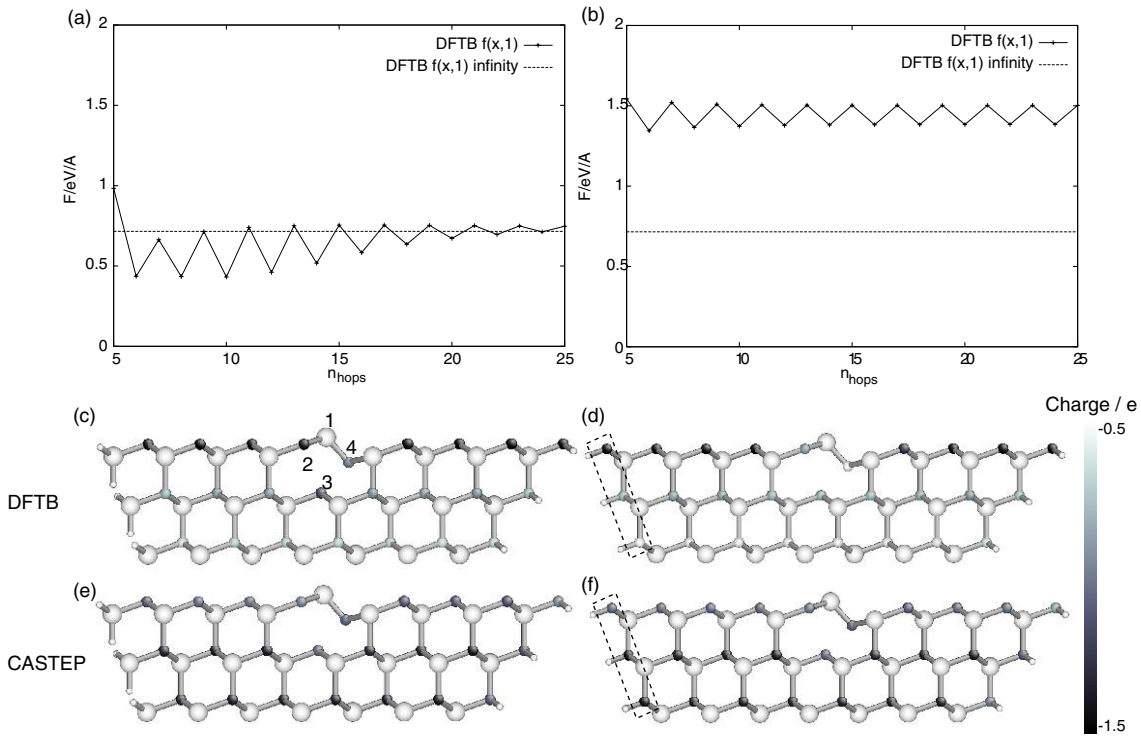

Figure 3. Example of the effect of long-range charge transfer on QM forces on a perturbed atom at the surface of a silicon carbide slab. The cell is periodic in the direction into the plane of the paper. (*a*) *x*-component of force on the atom labelled 1 in panel (*c*), calculated using DFTB for various sizes of unmodified clusters, together with the infinite periodic result. (*b*) Forces for the modified cluster, which includes three extra carbon atoms to improve the termination (shown in the dashed box in panels (*d*) and (*f*)). Note how in this case the DFTB calculations do not converge at all. (*c*)–(*f*) Atomic charges for a cluster surrounding the defect. (*c*) Unmodified cluster formed with $n_{\text{hops}} = 8$. Charges are calculated using DFTB. (*d*) Modified cluster, calculated using DFTB. The charges of the atoms labelled 2, 3 and 4 change markedly when the small change is made to the cluster surface. (*e*) Unmodified cluster using DFT (CASTEP). (*f*) Modified cluster using DFT (CASTEP). Note that with DFT there is no significant charge difference.

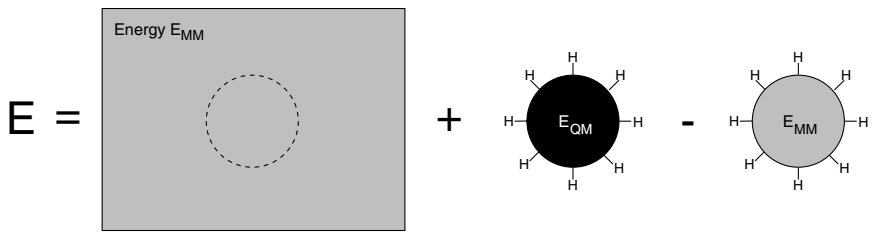

Figure 4. Schematic overview of the ONIOM method. The total energy is formed from three contributions: the MM energy of the entire system, and the QM and MM energies of the cluster.

the passivation atoms in the cluster calculations contributing to forces on the analogous atoms in the full system with a scale factor. For this subtractive method to eliminate boundary effects a delicate cancellation of errors must happen. The simplest example is for a perfect bulk system, where the true forces on all atoms are zero. The passivated cluster forces computed with QM and MM will in general be non-zero on the passivation atoms and nearby atoms. For these forces to cancel and lead to zero total forces in the coupled simulation, the passivated cluster forces computed with MM must be exactly equal to the QM cluster forces. In typical solid-state systems, this is not the case. For example, an unpassivated surface cut from a perfect bulk will have zero forces using the popular SW potential for silicon. Physically reasonable Si–H interactions might lead to non-zero forces on H-passivated surface atoms, but atoms deeper in the cluster will still have zero force on them. In any QM description, however, forces several atomic layers into the passivated cluster will be non-zero [32, 33]. The cancellation will not be perfect and may indeed be quite bad.

Ogata has developed a modified version of the ONIOM technique that makes it possible to add a buffer layer and lessens some of the above problems; this is illustrated in figure 5. The QM and MM clusters are surrounded by a buffer layer and an incomplete passivation of dangling bonds (see a more detailed discussion in section 5.3). The positions of the buffer and passivation atoms are relaxed using the MM potential while keeping the core cluster atoms fixed. The QM energy is evaluated on the cluster using the core atom positions from the coupled system and the buffer and passivation atom positions from the relaxed MM cluster. This makes the energy expressions for the QM and MM calculated clusters independent of the buffer and passivation atom positions.

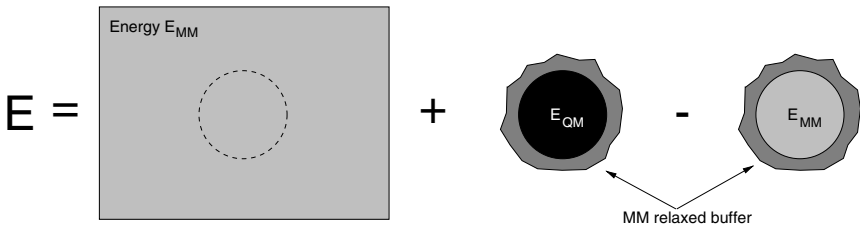

Figure 5. Schematic overview of the buffered-cluster approach. The buffer region is relaxed using the MM potential.

Therefore, the forces on the MM atoms near the QM region are exactly the full system MM forces. However, the forces on the QM region atoms are now the QM forces in the presence of atoms that are at positions different from the actual positions of the atoms in the full system. In the same way that ONIOM depends on the cancellation of errors due to the QM and MM descriptions of the cluster surface, Ogata's method depends on the cancellation of errors in the QM and MM descriptions of the *relaxed* cluster surface.

The ONIOM method and its various derivatives are 'subtractive' schemes. Additive methods attempt to correct for boundary effects in a different way [9, 30]. The energy is explicitly written as a sum

$$E_{\text{tot}} = E^{\text{MM}}_{\text{MM region}} + E^{\text{QM}}_{\text{QM region}} + E_{\text{coupling}},$$

where $E_{\text{MM region}}$ is the energy of the MM region, $E_{\text{QM region}}$ is the energy of the QM region and E_{coupling} accounts for the coupling between the two, and the superscripts again refer to which model is being used. In this approach essentially all of the sophistication is in the coupling energy, and the details are entirely dependent on the type of interactions present in the simulated system. Two examples are discussed in some detail below: MAAD, by Abraham *et al*, and the QUASI project. Both attempt to explicitly account for particular contributions to the interactions between the atoms.

4.2.2. Long-range interaction. Long-range forces, for example Coulomb interactions, present an additional set of challenges for seamless coupling beyond those of short-range interactions. If there is significant electrostatic coupling between the two regions, the QM and MM calculations cannot be carried out independently of each other: the QM calculation needs to include the field due to the MM atoms and vice versa. The only work to include this for a solid-state system was done in the QUASI project [30]. The QM calculation is carried out using not only the usual terms in the Hamiltonian for the QM atoms but also including in the QM external potential the electric field of the MM atoms. This interaction leads to contributions to the forces on distant MM atoms from derivatives of the QM energy. The remaining term is the classical electrostatic interaction between the QM nuclei and MM charges.

Placing classical point charges adjacent to the QM region when covalent bonds are broken at the QM/MM interface can lead to QM charge spilling into the MM region [30]. Several approaches can be used to deal with this boundary effect: ignoring potential matrix elements between basis functions on QM link atoms, ignoring potential terms from adjacent neutral MM groups in the QM Hamiltonian, shifting charge from the adjacent MM atom to nearby MM atoms when computing the QM Hamiltonian and representing the MM point charges as Gaussian distributions. More sophisticated approaches, such as specialized pseudopotentials [26] developed in the quantum chemistry context, have not yet been tried in the solid state.

5. Methods and applications

A number of groups have implemented various QM/MM methods for solid-state systems and applied them to various problems. By and large each group has addressed the subset of the above issues they considered to be relevant for their particular application. With two exceptions (namely, metal surface chemistry and defect chemistry in ionic materials), the applications have centred on the quintessential strongly coupled concurrent multiscale problem, the mechanical properties of systems with a crack or a dislocation, sometimes in the presence of chemical disorder. However, as the results show, not all of the test systems actually show the need for a coupled approach. Here we discuss each major scheme, summarize the scientific outcomes and try to put the method, tests and applications in the context of the issues we presented above. We recall with some detail the simulation setup of each application with the intention of recording a snapshot of current practices, because of the large amount of heuristics inherent in using the QM/MM technique.

5.1. Force mixing for fracture

The earliest QM/MM hybrid modelling of a solid system was done by Spence *et al* [19] (almost twenty years after the QM/MM approach was introduced by Warshel in the biochemistry community [34]). The scheme is introduced as 'flexible boundary conditions' for quantum mechanics and is basically abrupt force mixing but without any buffer region. The authors compute the equilibrium geometry and the lattice trapping barrier for a (1 1 1) crack in silicon. The quantum model is DFT for the inner 120 atoms. The MM region only contains 324 atoms, and the atoms in the outermost layer of the MM region are constrained to positions obtained from a continuum linear elasticity solution of the displacement field. The optimized geometries are obtained by alternate relaxation of the QM and MM atoms, always with the other set held fixed. The authors find a significant lattice trapping barrier to the propagation of the crack. Similarly to a later simulation by different authors [35], the size of the overall atomistic region

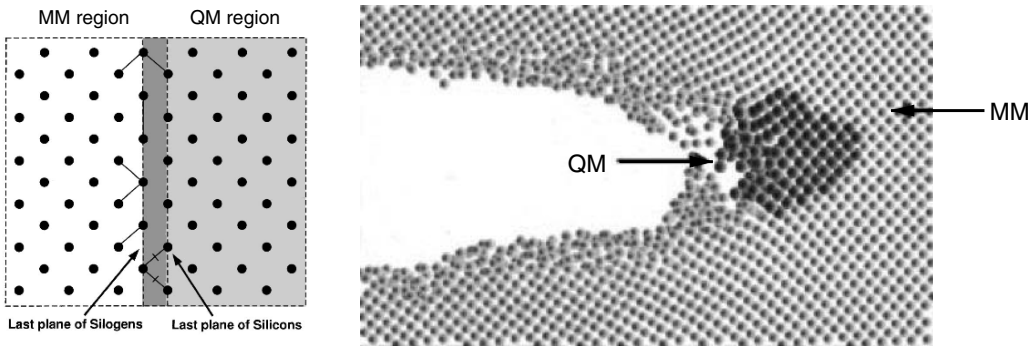

Figure 6. The MAAD scheme. (*a*) Schematic showing QM/MM handshaking. (*b*) Definition of QM and MM regions for silicon fracture simulation. Note the unphysically rough fracture surface. Reprinted with permission from [9]. Copyright 1999 by the American Physical Society.

was not large enough to properly account for the relaxation of the atoms near crack tip. As we discuss in section 5.9, the equilibrium structure of the silicon crack tip at near-critical loads is surprisingly different.

Five years later, the 'multiscale atomistic *ab initio* dynamics' (MAAD) method combined three computational descriptions: quantum mechanics, interatomic potentials and continuum elasticity using finite elements [9]. The coupling between the QM and the MM regions was originally formulated by defining a total energy that describes the combined system, although in the practical implementation an alternative formulation was used. The boundary between the QM and the MM regions is placed at a particular set of atoms. The QM region energy is computed by cutting out a cluster, including all interior atoms and the boundary atoms. Boundary atoms that are missing neighbours (from the MM region) are replaced by 'silogens', monovalent hydrogen-like atoms that are designed to give their neighbouring silicon a silicon-like bonding environment. The precise formulation of the silogens is tied to the tight-binding description used in the QM region: the silogens have a single s-orbital each, with on-site energy equal to the energy of an Si sp^3 hybrid, and Hamiltonian and overlap matrix elements with the Si s and p orbitals that are designed to mimic the interaction of two Si sp^3 hybrids. The silogens do not interact with each other. The energy of this cluster is added to the MM energy, computed using the SW potential, excluding pair and triplet terms that include only TB or boundary atoms. The forces on each atom are therefore simply the sum of the QM and MM forces, with boundary atom forces coming from a sum of silogen forces and modified MM forces. This can be thought of as an additive coupling approach. Since the interatomic potential uses first-neighbour bond-length and bond-angle energy terms, the analysis leading to the inclusion of various energy terms assumes that the same decomposition applies to the QM description as well. The coupling energy is the term that cancels out double-counted interactions that occur because the silogens and their first neighbours on the QM side are present in both the MM and QM calculations. All MM interactions involving only silogens and their QM neighbours are removed by the coupling energy, since they are supposed to be accounted for by the QM silogen–silicon interaction. For the pair interactions this is straightforward, although the assumption that an atom with two SW pair-interaction bonds and two silicon-silogen bonds will show bulk-like behaviour is unproven. Analogous arguments are made about the bond-angle forces. It is clear that this attempt to eliminate double counting is unlikely to be very accurate, because it assumes all QM bonding is essentially first neighbour and can be broken down into bond length and bond angle terms as in the MM.

Because the size of the QM region needed to describe the crack tip was too large for the desired computational speed, in the application the QM region was broken into overlapping subregions, computed independently. The QM forces were then computed using a type of abrupt force mixing: atoms that occur as QM atoms in multiple subregions follow forces that are the average of the subregion forces, while atoms that occur as QM atoms in one region and silogen or MM atoms in the other regions follow the QM force. The authors also propose a scheme for matching the properties of the MM and QM regions, although they do not implement it. In particular, they suggest matching the 'Einstein oscillator' force constants, i.e. the forces as a function of displacement of one atom in a fixed, bulk system. There is almost no explicit testing of the method. The only discussion is the empirical observation that, despite the use of force mixing, atomic kinetic energies are 'well behaved'.

The MAAD scheme was applied to dynamic fracture in silicon, as illustrated in figure 6. A periodic thin (10.9 Å) Si slab with a thin crack exposing (1 0 0) faces was placed under uniaxial tension imposed by initializing a velocity gradient along the pulling direction and constraining the outermost surfaces to move at constant speed. The strain rate was reduced to zero (no details given as to how exactly) once the crack began to propagate. The overall system was about 3650 Å (periodic) along the crack direction, and about 5600 Å along the pulling direction (including the finite-element region surrounding the atomistic region). The initial crack was not specified, but appears from the figure to be about 30% of the 3650 Å length. One crack tip was surrounded by a QM region, while the other was not. Crack propagation was observed at a tensile strain of about 1.5%, with the crack rapidly accelerating to 85% of the Rayleigh wave speed. There was no difference between the behaviour of the two crack tips in the crack speed as a

function of time, and the QM crack tip appears to be disordered. The authors note that the crack surface is initially smooth (on larger than atomic length scales) and gradually roughens. They speculate that this roughening is related to Rice's proposal of sessile dislocations that are trapped at the crack tip [36], which however was later found to be based on an inaccurate estimate for the Si unstable stacking fault energy [37].

As later experiments [38] and simulations [7, 10] showed, silicon is in fact brittle, and the extensive (if localized) crack-tip plasticity in the MAAD simulations is unphysical. Although there are several potential sources of error, it is not clear why they lead to such qualitative failure. The first possible issue is the formulation of the silogens. As the paper states, the parameters for the silogen–silicon interaction are designed to mimic the interaction of a bulk silicon atom with another similar atom. However, the role of a surface termination atom is more like mimicking the interaction of the cluster Si with a semi-infinite slab of bulk Si. It is not clear that the silogens do this effectively. It is also unclear whether the silogen–silicon bond energies plus the modified SW bond energies (excluding certain interactions as described above) reproduce the energy of a bulk Si atom. The second issue is the breakup of the QM region into overlapping subregions, in this case cylinders with radius 5.43 Å. It is known that in finite Si clusters carved from bulk, whether terminated by H or by optimized orbitals [32], atoms that are several bond hops from the cluster boundary can have qualitatively incorrect forces [33] compared with the original bulk. Given the subregion radius used, most of the atoms in each subregion followed trajectories generated by substantially incorrect forces.

5.2. *FE/MD/EDFT*

The hybrid finite-element/MD/electronic-density-functional-theory scheme presented by Ogata *et al* also combines three computational descriptions of the material [24]. The QM–MM coupling is achieved using a straightforward ONIOM-type approach: the total energy is expressed as the energy of the system using the MM potential, plus the energy of an H-terminated cluster using DFT, minus the energy of a Si-terminated cluster using the MM potential, as in equation (2). The terminating H positions lie along the broken Si–Si bonds, at a distance proportional to the Si–Si distance. The value of the proportionality constant is optimized to minimize the forces on the cluster atoms (in an unspecified geometry). The implementation uses real-space multigrid DFT with norm conserving pseudopotentials and the SW potential for the MM. Matching the properties of the two descriptions is not discussed.

The method has been applied to O_2 dissociation on the Si(1 1 1) surface [24] and the (1 0 0) surface [39] and the effect of H_2O on Si (1 1 0)-surface crack initiation [40]. There are no tests for the accuracy of the coupling method.

The O_2 dissociation on the Si (1 1 1) surface [24] was carried out in a 213 Å × 245 Å wide, 83 Å thick slab, where the top 12 atomic layers are treated atomistically. The QM cluster included 10 Si atoms, 13 terminating H atoms and the O_2 molecule. The molecule is initially placed parallel to the surface along the $[\bar{2}\,1\,1]$ direction, and the system is propagated in time with MD for about 900 fs with a 1 fs time step. The molecule dissociates, and each O atom is incorporated into an Si–Si bond. The authors observe that the kinetic energy released by the dissociation propagates into the MM region within 300 fs and later into the finite-element region. The total energy is conserved, fluctuating by about 10^{-6} eV. A similar simulation of O_2 dissociation on the Si (1 0 0) surface [39], using only QM and MM regions, was carried out on a 65 Å × 65 Å wide, 35 Å thick slab, with the O_2 molecule initially aligned normal to the surface. The system was propagated in time using MD with a 1 fs time step. The QM clusters included 20 Si atoms and 15 passivating H atoms, as well as the molecule. The molecule first tilts so that it is parallel to the surface and then dissociates over about 250 fs. One O atom is incorporated into an Si–Si dimer in the QM region centre, and the other approaches the edge of the QM region. The authors show that the energy released, which locally increases the temperature over 1000 K, propagates into the MM region and that the total energy is conserved up to fluctuations of about 10^{-4} eV. This 'hot atom' dissociation mechanism has later been reported using first principles calculations [41], although the local temperature only reached ∼550 K in that case.

The H_2O on Si (1 1 0)-surface crack simulations were carried out on a system that is 68 Å thick along the crack propagation direction, 46 Å thick along the crack front direction and 153 Å thick along the pulling direction. A 34 Å notch was created by removing one atomic layer normal to a (0 0 1) surface. Three QM regions, each surrounding a single H_2O molecule, are defined at the crack-tip region, where there are Si dangling bonds from the crack-surface atoms. Each region includes 36 Si atoms, 40 termination H atoms and the 3-atom water molecule. The system is relaxed (with the QM regions but without the water molecules) at 3% and 4% uniaxial tension, corresponding to stress intensity factors (presumably defined using the stress in the MM region) of about 0.4 MPa$\sqrt{\text{m}}$ and 0.5 MPa$\sqrt{\text{m}}$, respectively (the experimental K_c is reported to be about 0.9 MPa$\sqrt{\text{m}}$). The H_2O molecules are added at random orientations, and the system is propagated in time with 1 fs time steps for 500 fs. At a low load, two of the three molecules dissociate on a 100 fs time scale into an H atom and an OH group, each bonded to previously dangling Si bonds. The third molecule dissociates into an O atom that inserts itself into an Si–Si bond and two H atoms bonded to dangling bonds. At a higher load, two of the three molecules dissociate with an O atom incorporated into an Si–Si bond. For the third molecule the crack tip advances by the breaking of two Si–Si bonds. The water molecule itself dissociates into an H atom bonded to a preexisting Si dangling bond and an OH group bonded to a newly created dangling bond of the Si atom previously at the crack tip. The authors also analyse the charge transfer that occurs during the reaction processes. The brief duration of the simulation and the small number of molecules involved make it difficult to determine whether the observed behaviours are typical and whether the differences between the two loading conditions are systematic. A similar but shorter (130 fs) simulation for a crack under

3% strain where the dangling bonds near the crack tip were saturated with H [42] showed similar results to the unsaturated crack surface. In this case, however, two of the three molecules remained unchanged, while the third dissociated into an O atom, which was incorporated into a strained Si–Si bond, and two H atoms with unclear bonding to the four-fold coordinated crack-tip Si atoms.

5.3. Relaxed buffered ONIOM

The 'buffered-cluster' method of Ogata is a modification of the conventional ONIOM technique used in the FE/MD/EDFT method [25]. Ogata notes that the ONIOM method is sensitive to the particular choice of QM atoms and that 'inappropriate' choices lead to significant distortions of the atoms near the boundary. The only specific reason given for this is that H atoms that are close lead to a repulsive interaction. Ogata also notes that for systems with less obvious covalent bonds, e.g. α-Al_2O_3, H atoms do not result in good passivation. To address these issues Ogata modifies the prescription for passivating the QM and MM clusters.

In the silicon case the MM cluster is passivated by a single layer of Si atoms whose positions are varied so as to minimize the MM energy of the cluster while holding the core atoms fixed. The QM cluster is passivated by a layer of atoms, either H or Si as described below, at positions corresponding to the *minimized* MM cluster passivation atoms. Passivation sites that saturate one broken bond are populated by an H atom at a scaled distance along the bond. Sites that saturate two or three broken bonds are populated by a Si atom. Because the positions of the passivation atoms in the MM cluster minimize the energy, there is no contribution to the forces from derivatives with respect to those atoms.

To apply the method to α-Al_2O_3, the cluster surface is restricted to being composed entirely of Al atoms. The passivation region (where positions are relaxed to minimize the MM energy) consists of a layer of O atoms and a layer of termination atoms. In the MM cluster the second layer consists of Al atoms, while in the QM cluster it consists of either Al (for sites that passivate more than one Al–O bond) or a $0.5e$ core charge (for sites that passivate one broken Al–O bond). There is no Coulomb interaction between the QM and MM clusters and the rest of the system: such terms are calculated entirely from the Coulomb term in the full system MM calculation.

Some basic tests of the method are presented for silicon. For some small clusters (1–29 atoms) cut out of the bulk structure, the maximum displacement of the atoms in the relaxed structure away from the diamond structure is 0.03 Å. The Einstein oscillator force constants, i.e. the force as a function of displacement for a single atom in a fixed lattice, are plotted for atoms in the centre of the cluster, at the QM/MM boundary and in the MM region, displaced along high symmetry directions. For small displacements, less than 0.1 Å, restoring forces on the different displaced atoms are within a few per cent of each other (there is no comparison with full QM calculation), and deviations grow up to about 20% for displacements around 0.4 Å. The passivation Si atoms (used to passivate 2 or 3 dangling bonds) lead to surface states, and Ogata shows that they are well localized, although there is no discussion on whether there could be charge transfer when bonds break in the QM region core. Instead, it is shown that changing the total charge on the cluster leads to negligible changes in atomic forces (suggesting that the test is not relevant, since clearly there are physically interesting configurations during bond breaking where the forces are dependent on the amount of charge transfer, as shown, for example, in our SiC test in section 4.2.1). It is noted in passing that the passivation Si dangling bonds could in fact be passivated with more H atoms, as well. A similar accuracy is seen for the α-Al_2O_3 bulk.

Ogata also gives a procedure for updating the QM region at intervals of 10 fs. Atoms near the dissociating O_2, with the coordination number greater than 4 or with coordination number 4 but at least one neighbour at more than 1.2 times the equilibrium bond length, become QM atoms. Also, all atoms that are on the shortest path between some pair of QM atoms become QM atoms. Finally, atoms at the edge of the QM region that have 4 bonds within 10% of the equilibrium bond length switch back to MM. The method is used to study the effect of water molecules on a crack tip [42] and the dissociation of a water molecule at the tip of a crack [25].

In the first simulation, the buffered-cluster method is applied to the simulation of a blunt crack tip in α-Al_2O_3 in the presence of water. A slab with a [0 0 0 1] surface, about 84 Å wide in the [1 $\bar{1}$ 0 0] direction and about 24 Å thick, is initially cracked by a thick planar notch with rectangular cross-section, normal to the [0 0 0 1] surface and aligned along the [1 1 0 0] direction, reaching about half-way through the slab thickness. The unstrained system and a system under 3% uniaxial tension normal to the notch surface are relaxed, and an H_2O molecule is placed near the blunt notch tip. In the unstrained system the water is adsorbed. Under tension the molecule is adsorbed to a surface Al over 40 fs, and at around 60 fs an Al–O bond breaks, and the crack propagates (in a 'dry' manner, without water activity) for over 240 fs until it reaches the bottom of the slab.

More detailed simulations [25] were carried out on water molecules in notched strained Si slabs. A system with [0 0 1] surfaces, 109 Å × 71 Å wide and 27 Å thick, is pre-cracked by a wide nearly rectangular cross-section planar notch normal to the surface and forming [1 $\bar{1}$ 0] surfaces, with small [1 1 1] facets near the blunt tip. Two different notched systems are created, with the notches differing in depth by one atomic layer, leading to [1 0 0] surface reconstruction dimers at the blunt crack tip aligned along two orthogonal directions. The system is placed under uniaxial tensile strain of 4% normal to the crack surface, as illustrated in figure 7.

The system is relaxed with the MM potential, and the virial stress is measured for the two systems with different depth notches. The stress is about 2 GPa far from the crack, about 8 GPa maximum near the crack tip, and the stress field is qualitatively different depending on the reconstruction dimer orientation. A QM region is defined at the centre of the crack tip, and the hybrid simulation is applied to the one unstrained notch and the two different notches for the strained system. The reconstruction dimers buckle in the unstrained system and

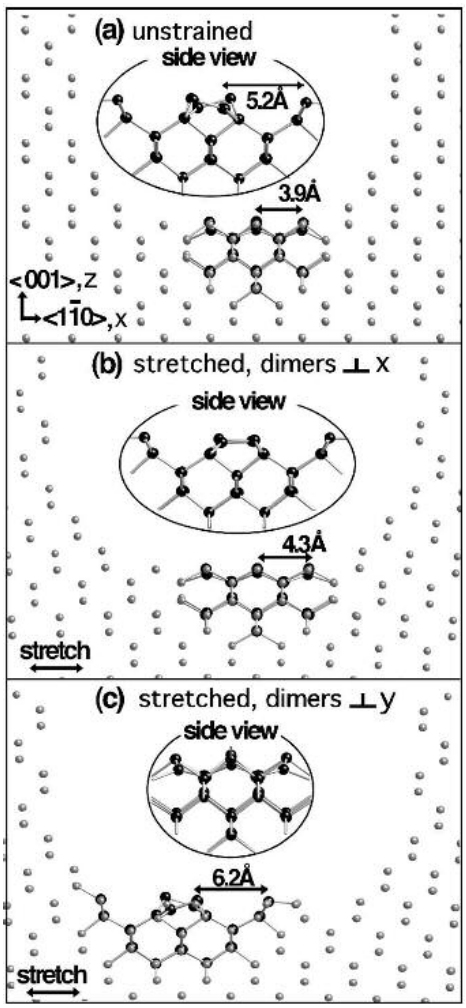

Figure 7. Application of the relaxed buffered ONIOM scheme to a notched Si system. The QM Si atoms are shown in black. The three panels show an unstrained system and two stretched systems. Note that the [1 0 0] surface reconstruction dimers in panels (*b*) and (*c*) are orthogonal due the different depth notches. Reprinted with permission from [25]. Copyright 2005 by the American Physical Society.

in the system where the dimers are parallel to the stretching direction, while the dimers in the perpendicular orientation buckle very little.

An H_2O molecule is added, and the system is relaxed while constraining the Si–O distance. Two different cluster sizes, 33 Si atoms and 61 Si atoms, are compared, and since the force differs by only 5% the authors use the smaller cluster for further calculations. An adsorption energy of 0.7 eV for the unstrained system, with no barrier, is calculated by computing the total energy as a function of the Si–O distance, in agreement with the published values [43, 44]. The strained system shows about half the adsorption energy, and only the strained system with buckled dimers shows no barrier to adsorption; the system where the strain suppresses buckling shows a small barrier to adsorption along this particular reaction coordinate. The authors assert that this is because the buckling is necessary for the adsorption to occur.

The authors then carry out calculations of some pathways for dissociation of the water molecule into an OH group and an H atom by constraining the Si–O and Si–H distances, forcing the molecule to dissociate onto dangling bonds on the same dimer, nearest-neighbour dimers (along the dimer row) and next-nearest-neighbour dimers (across the trench between dimer rows). For the unstrained system, which is equivalent to a flat [1 0 0] surface, they find barriers of 0.15–0.5 eV and dissociation energies of 1.8–1.5 eV, respectively, depending on the final geometry, in agreement with the published values [43]. For the strained system they find barriers of 0.05–0.6 eV and energies of 1.6–0.5 eV for the buckled dimer configuration and barriers of 0.1–0.8 eV and energies of 1.7–0.5 eV for the flat dimer configuration. Since the crack tip is flat and the water molecule is placed far from the corners, the crack tip is equivalent to a flat surface at a larger strain than that applied to the full system (due to the crack-tip stress enhancement), and the function of the coupling is merely to produce the appropriate static stress enhancement.

The authors note that the dynamics of oxidation of the Si surface depends on whether dimers typically have species saturating both dangling bonds or if one dangling bond is left unsaturated, and therefore the statistics of the dissociation paths of the water molecule are important. The authors simulate this by setting up a relaxed system, and running five MD simulations with a water molecule, all three atoms initialized to the same velocity equivalent to $T = 600$ K, moving in five different directions. The MD time step is 0.67 fs, and the QM region is recomputed every 10 fs as described above. In two instances the molecule dissociates onto next-nearest-neighbour dimers, and in the other three the molecule does not dissociate. There is no explanation for why this pathway, which was least favourable energetically, dominates. The authors also note that there is no effect from the notch side walls; this is similar to the case above in this respect. In the strained system all five initial conditions lead to dissociation, but in this case the H and OH groups bond to dangling bonds on the same dimer.

5.4. DCET

The dynamical coupling of empirical potentials and the tight-binding (DCET) method uses simple abrupt force mixing [10, 45]. The system is divided into QM and MM regions. The QM forces are computed in a cluster buffered by a layer of atoms defined by a distance cutoff. Atoms outside the QM region, including those in the buffer region, follow trajectories generated by MM forces. The QM forces are calculated using the linear-scaling GF based method described in section 4.2.1, which passivates the QM cluster calculation by applying constraints on the solution to the electronic problem in the buffer region.

The method is applied to simulations of fracture in Si. The MM potential used is the environment dependent interatomic potential (EDIP) [46], modified by scaling the EDIP length scale to match the lattice constant to that of the tight-binding model used in the QM region. The use of a TB model also facilitates tabulation of the perfect bulk GF matrix elements at the beginning of the simulation. The buffer region consisted of a 6.5 Å thick layer adjacent to the QM

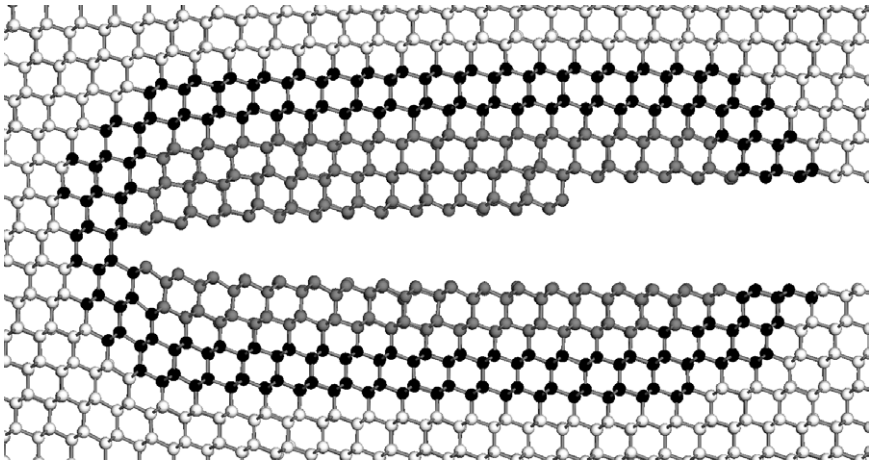

Figure 8. Snapshot 2 ps into a DCET simulation of crack propagation on the silicon (1 1 1) cleavage plane. White signifies atoms simulated using the MM potential, grey signifies atoms in the tight-binding (QM) region and black signifies buffer atoms at the interface between the two regions. Reprinted with permission from [10]. Copyright 2003 by the American Physical Society.

region. There are no explicit tests of the method published, either in terms of boundary effects or energy conservation of the force-mixing approximation. However, it was verified that energy conservation is approximately obeyed by monitoring the kinetic energy of the entire system, 54 000 atoms, during MD simulations over periods of about 10 ps. Since the overall system size is much larger than the QM region, and any spurious heating would be localized to the QM/MM boundary, this test is actually not a particularly sensitive test of energy conservation. The criteria for selecting the QM region, which were not published, consisted of a predefined geometrical region (stadium shape) placed at a particular position relative to the crack-tip atom, defined as the extremal (along the crack propagation direction) atom with coordination less than 4.

The authors simulate fracture in a thin periodic Si slab (see figure 8), about 12 Å thick and 235 Å × 400 Å wide, with a (1 1 1) surface notch, created by removing an atomic bilayer, extending for half the longest dimension of the sample. The system is placed under uniaxial tension normal to the (1 1 1) notch surface by constraining the outermost few atomic layers of the system. A TB region about 17 Å wide and 55 Å long, surrounded by a buffer region, is fixed during the simulation. The system is initialized at various strains, relaxed using the MM potential only, and then propagated in time using velocity-Verlet MD with a 1 fs time step. The crack speed is measured in a range of applied strains. The fracture process is brittle, with no disorder in addition to the opening of the crack surface. The simulation shows that crack propagation begins once the energy release rate exceeds two times the QM surface energy, i.e. nearly at the Griffith critical load, and this is confirmed by quasistatic energy relaxation. However, the crack-tip reconstruction observed with other TB methods and DFT calculations [7] was not observed. It is unclear whether this failure is attributed to the TB model [47], the large electronic temperature implied by the small number of approximate Fermi function poles [10] or some other factor.

The authors attempt to compute the energy barrier to crack propagation (lattice trapping barrier) despite the lack of a total energy in the method. They do this by computing the barrier for the MM system and subtracting from it the MM energy as a function of separation of two slabs with a flat (1 1 1) surface. The remaining energy is attributed to elastic relaxation, which is assumed to be the same as for the coupled simulation. This energy is then added to the QM slab separation energy, and the sum shows a small barrier at the Griffith critical load, which goes to zero at the load where crack propagation is observed. They conclude that this decomposition into elastic-relaxation and bond-breaking energies is physically meaningful and that the relevant parameters are therefore the energy of the separating slabs and the shape of the crack tip which controls how quickly the elastic energy is relieved.

5.5. GF based additive coupling

Nieminen and Paavilainen have developed a formalism [48] for coupling a locally charge-neutral tight-binding model and an interatomic potential of Finnis–Sinclair form [49, 50] for transition metals based on a GF expression for the density matrix. The TB problem is solved in a space of basis functions centred on QM atoms and on atoms in a buffer region composed of nearby MM atoms. The pair repulsion contributions to the TB energy are not discussed below, as they simply correspond to an additional MM-like term. All remaining energy contributions are written in terms of the trace of the product of a density matrix and a Hamiltonian matrix, $\mathrm{Tr}\rho H$. In the MM region, the form of the interatomic potential makes it possible to transform the usual Finnis–Sinclair energy expression into an effective density matrix and an effective Hamiltonian matrix. The rest of the energy is computed from the TB model Hamiltonian and a density matrix computed as a line integral of a GF matrix over complex energy. The GF itself, G, is built up from an initial value G^0 by solving a Dyson equation

$$G = G^0 + G^0 V G,$$

where V is the off-diagonal part of the Hamiltonian matrix. This process is carried out iteratively, repeatedly solving the equation for 2×2 blocks of V. The initial guess for GF matrix elements between basis functions on QM atoms is generated exactly from the diagonal TB Hamiltonian for free atoms. For

GF matrix elements between buffer atoms, however, the initial guess is computed from analytical approximations to diagonal blocks of the GF appropriate for bulk transition metals [51, 52], which are environment dependent. The authors note that their total energy expression neglects the contribution from QM atoms to the density used to compute interactions between MM atoms, but they state that this effect is minor. The forces are claimed to be the derivatives of the total energy, and the Hellmann–Feynman [53, 54] theorem is used to justify the neglect of contributions from derivatives of the density matrix with respect to atomic positions. When the correct GF is used to construct the density matrix, this reasoning holds. However, the Dyson equation yields the correct Green's function G from the initial guess G^0 only if the off-diagonal potential V is the difference between the full Hamiltonian H and *the Hamiltonian* H^0 *that yields* G^0. Since the elements of G^0 in the buffer regions are environment dependent, they depend on nearby atomic positions. Therefore, the corresponding Hamiltonian is atomic position dependent. However, these diagonal elements are not included in the total energy, and their derivatives do not contribute to the atomic forces. With these terms neglected it does not appear that the Hellmann–Feynman theorem applies. It is unclear whether there is also double counting implicit in combining a GF initial guess for the buffer atoms derived for a bulk system with all the off-diagonal matrix elements connecting QM and buffer atoms; presumably some of those interactions were already accounted for by the use of diagonal blocks from a bulk GF.

The method was used to evaluate forces for a MD simulation of the approach and dissociation of an O_2 molecule on the Pd(1 1 0) and (3 2 0) surfaces. The MD simulation was carried out using the Nosé–Hoover thermostat. The system consisted of a finite slab, periodic in plane with 64 Pd substrate atoms for the (1 1 0) surface and 192 atoms for the (3 2 0) surface. There is no discussion on the tests of the method, nor how the QM region was selected, and the buffer region is described as consisting of atoms within a 'sufficient cutoff radius' around the QM atoms. Energy conservation during dynamics (which may show the effect of inconsistent forces and energies) is not discussed either. Initially the entire Pd substrate is described as being treated with the interatomic potential, but the number of orbitals mentioned in the context of computational efficiency is consistent with 6–8 Pd QM atoms near the beginning of the simulation and about 20 after dissociation (assuming that the two O atoms are always in the QM region). Clearly some algorithm must have been used to determine the QM region dynamically during the simulation.

The authors find that on the unreconstructed (1 1 0) surface the O_2 molecule has multiple minima, but with a missing row reconstruction (suggested by the experiment) it always binds to hollow sites on (111) facets, in agreement with the experiment [55]. They also find that the π-bond contribution to the intramolecular bond decreases gradually as the molecule approaches the surface, while the σ contribution decreases abruptly only once the molecule has adsorbed and the two O atoms move apart. The clean (3 2 0) surface remains unreconstructed in the simulation, but the dissociating molecule causes significant rebonding. The need for coupling

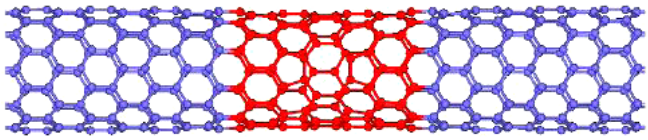

Figure 9. Application of ONIOM to CNTs. The QM region is shown in red and the MM region in blue. Reprinted with permission from [11]. Copyright 2007 by the American Physical Society.

emerges in this system, because the clean surface structure is not even stable until the slab is more than eight layers thick, requiring a number of atoms that is unfeasible for a fully QM calculation. The O atoms become buried, and the surface steps inherent to the unreconstructed surface become less distinct. The authors cite a private communication ([23] in [48]) of an experiment suggesting confirmation of reconstruction only in the presence of adsorbates, consistent with the simulation, although a later publication [56] discusses the reconstruction without mentioning the adsorbed species.

5.6. ONIOM for nanotubes

An ONIOM scheme was used to simulate carbon nanotubes (CNTs) [11] (see figure 9), which are in some sense between a conventional chemical system and a solid-state material. The Tersoff potential was used in the MM region and two different models, PM3 [57] and local-orbital DFT (SIESTA [58]), were used in the QM region. The cluster was passivated with H atoms positioned along broken C–C bonds at a fixed distance, 1.09 Å, from the neighbouring C atom. To improve the matching of properties, the authors scale the MM energy to match Young's modulus of the appropriate QM model. They note that without this scaling the MM potential is significantly less stiff than either QM model, and fracture could occur in the MM region at unphysically low loads.

There are no specific tests of the model for force accuracy or cluster edge effects, but comparisons are made with full QM calculations for the loading of some small CNTs ([5, 5] and [10, 0]). CNTs with single vacancies, divacancies and Stone–Wales defects were subjected to tensile strain until failure. Below failure the authors found good agreement for the stress versus strain curves and substantial deviation from a pure MM calculation. At higher loads the coupled simulation stresses were 3–5% higher than the full QM simulations. The failure stress of the coupled simulation was also in good agreement with the full QM calculation, although the failure strain was about 10–20% lower.

The authors note that finite length CNTs show Peierls distortions that can lead to multiple fracture pathways and that the coupled method will not reproduce those (the MM cannot show the Peierls distortion, and the QM region may show it but is different in detail from the full QM CNT). They also note that when the QM region is a cylindrical section of a CNT, rather than a patch surrounding a defect, the coupled method converges more slowly to the long tube limit, although they do not speculate about the cause.

Calculations were then carried out on much larger (50, 0) and (29, 29) tubes, containing about 4000 atoms and about 90 Å long. The fracture stresses for CNTs with different size

hexagonal holes and with different length slits (normal to the tube axis) were computed. The authors also checked for the effect of the QM region size (for the smaller slits), and found no difference in the mechanical properties. They find that the fracture stress goes down with increasing defect size, and that the crack is normal to the tube axis as long as it is within the QM region. They also find that armchair tubes have higher fracture strengths than zigzag tubes, despite the fact that the cracks in the former have lower surface energies. They attribute this to the greater effect of relaxation on surface energies of armchair CNTs and suggest that this leads to higher lattice trapping as discussed by Pérez and Gumbsch for Si [35]. They analyse the details of bond breaking and show that the maximum strain of the crack-tip bond is higher for the coupled simulation than for the pure MM simulation, although both predict the same bond will fail first. Circular holes, which do not have sharp corners near the maximum stress concentration, show much larger fracture stresses. Comparisons with the stress implied by the Griffith criterion on similar slits in larger flat sheets show reasonable agreement, although the plot as a function of the slit length seems to have a systematic deviation from the Griffith prediction functional form. The fracture load was computed by applying the isotropic K-field displacement at the boundary of a region of unspecified size at increasingly larger values of K_I until failure and measuring the energy release rate by evaluating a discrete form of the Rice J-integral [59]. The critical loads show lattice trapping, in the form of critical energy release rates of 10–20% higher than the Griffith criterion prediction.

Other works have reported little success when applying the ONIOM method to nanotubes. Chen *et al* [60] report that treating nanotube systems modified by the addition of oxygen, methylene or NH with ONIOM does not reliably reproduce the full QM results. A study of chemisorption of ozone on nanotubes concludes that to get accurate results, the ONIOM QM region needs to be so big as to give no advantage over a full DFT calculation [61]. An application of the ONIOM method to functionalized carbon nanohorns [62] finds that ONIOM captures the essential trends, but significantly overestimates binding energies when compared with DFT. These observations provide the motivation for a recently proposed modification to the ONIOM technique in which the QM region is excluded from the MM calculation [63]. This method has been applied to the fracture of graphene sheets, giving fracture strengths in good agreement with a fully QM calculation.

5.7. ONIOM for metals

Choly and coauthors proposed two hybrid schemes in [12]. They start with the basic additive energy-mixing idea:

$$E^{\mathrm{QM/MM}}_{\mathrm{QM+MM}} = E^{\mathrm{MM}}_{\mathrm{MM}} + E^{\mathrm{QM}}_{\mathrm{QM}} + E^{\mathrm{interaction}}_{\mathrm{QM+MM}}, \tag{3}$$

where the superscripts denote the energy function and the subscripts the region over which it is evaluated. The first two terms represent the independently calculated MM and QM energies in their respective regions and the last term represents the 'interaction' energy, i.e. all the errors that one commits by computing the first two terms. If the interaction energy is computed using the MM model,

$$E^{\mathrm{interaction}}_{\mathrm{QM+MM}} = E^{\mathrm{MM}}_{\mathrm{MM+QM}} - E^{\mathrm{MM}}_{\mathrm{QM}} - E^{\mathrm{MM}}_{\mathrm{MM}}, \tag{4}$$

then the method is equivalent to the subtractive ONIOM scheme. On the other hand, if a QM method is used, then

$$E^{\mathrm{interaction}}_{\mathrm{QM+MM}} = E^{\mathrm{QM}}_{\mathrm{MM+QM}} - E^{\mathrm{QM}}_{\mathrm{QM}} - E^{\mathrm{QM}}_{\mathrm{MM}}, \tag{5}$$

which would seem like accomplishing nothing, except that the second term is cancelled by the corresponding term in equation (3) and the first and third terms can be calculated together taking advantage of the large degree of cancellation between them in the MM region in a purely real-space formulation of quantum mechanics. Indeed, this scheme can be classified as an energy-mixing scheme with a finite buffer region, where the buffer region is defined as that part of the MM region in which complete cancellation of the first and third terms in equation (5) is not assumed. In the context of DFT, this scheme is reliant on an orbital-free formulation of the kinetic energy functional and a real-space solution which also requires local pseudopotentials. Using such a formulation, the QM energy corresponding to an arbitrary charge density can be computed. By fixing the charge density in the MM region to a superposition of charge distributions centred on each atomic position, the calculation of the interaction energy only involves a computational effort close to that of calculating the QM energy of the QM region. If the atom-centred charge distributions in the MM region reproduced the QM charge density for a perfect crystal (cf bulk property matching), this scheme would have small force errors for the perfect crystal. The authors instead chose a simpler, spherically symmetric charge density and show that in this case the scheme has significant errors in the forces in a perfect block of fcc aluminium, up to 0.3 eV Å^{-1}. The large force errors reported for the ONIOM variant are not unexpected, given the poor description of the artificial surfaces in the MM model, as discussed above. Nevertheless, the reported force errors are much smaller than what would have been obtained for a covalent system, because in metallic systems there are no cluster termination issues, and the empirical atomistic model (EAM [64, 65] in this case) is a decent approximation to the DFT model.

A later paper by the same group combines the quasicontinuum method [15] with ONIOM and applies it to the problem of the core structure of an edge dislocation in aluminium. The authors performed unspecified convergence tests and settled on using an 84 atom cluster in their DFT calculations. The main result of the paper is that the splitting distance of the two partials into which the dislocation dissociates shrinks to 5.6 Å from the original value of 15.4 Å as computed with pure QC, i.e. without quantum mechanics. This smaller splitting distance is consistent with another report [66] and experiment [67]. In addition, the authors investigated the effect of H impurities on the dislocation core and found that H-containing cores are larger (about 13 Å) and there is significant charge buildup near the H atom.

Recently, Liu *et al* reported [68] a modification of the above schemes. They add a term to the total energy function

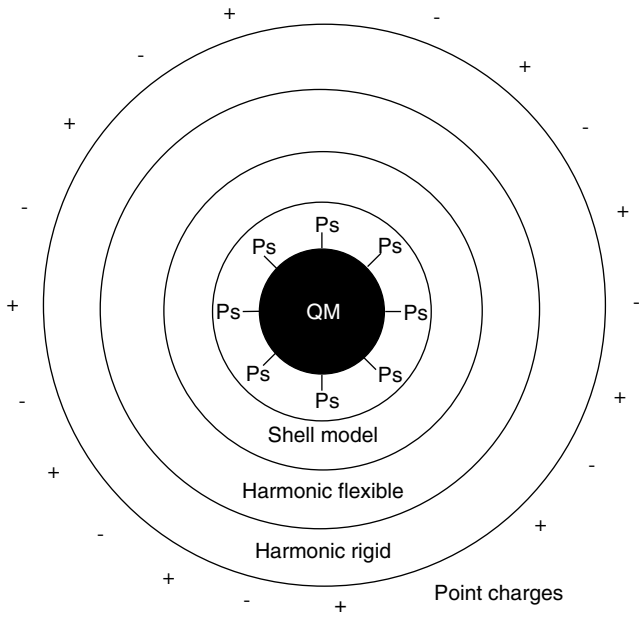

Figure 10. Schematic representation of the hybrid scheme for ionic systems. The QM cluster is terminated by pseudopotentials and surrounded by several layers of MM atoms which are described by a sequence of models decreasing in sophistication and computational cost: a shell model, a harmonic approximation to the shell model, the same harmonic approximation with fixed atom positions, and finally a set of point charges that are fit to reproduce the local field of the infinite crystal.

that results in forces equal to those from abrupt force mixing. The resulting total-energy-like quantity, defined so that the forces are its derivatives, is not equal to the physically meaningful energy. In particular, the new term is explicitly time dependent, so the dynamics does not conserve energy and requires a thermostat to keep it stable. Nevertheless, the force errors on a perfect crystal are much reduced as compared with the ONIOM scheme, as they should be, since the force-mixing scheme has a buffer region. They use a buffer region of one atomic layer around a QM region of 126 atoms and calculate the stacking fault width of an edge dislocation in aluminium to be about 5.9 Å and the Peierls stress to be 7.5×10^{-3} in units of the Al shear modulus, which is within the upper range of experimental results.

5.8. Ionic QM/MM

The QUASI project implements a QM/MM method designed for materials with significant ionic bonding [29, 30]. The scheme is now part of the ChemShell simulation package [69]. The coupling is an additive scheme, with the total energy written as a sum of the energy of the QM cluster, the MM system without the QM atoms and a coupling term, as illustrated in figure 10. The coupling between the regions is assumed to be dominated by electrostatic and pair-wise short-range interactions. A shell model is used for the MM region, and the MM energy consists of the shell displacement energy, plus short-range and long-range (Coulomb) interactions between MM atoms. A part of the QM–MM interaction energy is accounted for by the electrostatic potential of the QM nuclei which is applied to the MM atoms. The rest is accounted for by embedding the QM atoms in a potential related to the MM atoms.

The QM calculation includes the usual intra-cluster terms in the Hamiltonian. For coupling and passivation, the QM potential includes long-range electrostatic terms (centred around nuclei and around shells) and short-range pseudopotential terms (centred around nuclei) associated with MM atoms. The pseudopotentials are used to localize the electrons, correct for termination effects on the band gap, and compensate for missing short-range forces in the QM–MM interface. Conventional pseudopotentials are used for MM cations, and due to the lack of anionic pseudopotentials their (non-electrostatic) contribution to the QM potential is neglected. Since cation pseudopotentials do not have explicit semicore states, a short-range pair-wise interaction energy between QM cations and MM cations is added to the MM energy.

To compute defect energies, the MM region is further partitioned. Forces on inner MM atoms are computed from the full interatomic potential, while forces on outer MM atoms are computed from a harmonic approximation. In the inner MM region all atomic positions are relaxed, while in the outer MM region, only atoms near the inner region are relaxed, while atoms further away are held fixed. Finally, point charges on the outer surface of the atomistic region (a spherical shell for bulk or interface calculations or a hemispherical or cylindrical one for surface calculations) are placed to reproduce the electrostatic field in the inner MM region caused by the missing infinite solid. To improve the convergence of the electrostatic potential, the atomistic region's edge is modified to include entire compact neutral groups, preferably with small lower electrostatic multipoles. The scheme has been applied to electrons binding to O vacancies at the ZnO surfaces [29], formate on the ZnO surface [30] and N_2O adsorption on Cu-exchanged zeolites [30].

The embedding method was used to simulate defects (vacant oxygen interstitial surface sites) in the ZnO$(0\,0\,0\,\bar{1})$ polar O-terminated surface and their interactions with localized electron states, where periodic slab calculations using LDA have been done previously [29]. The system consists of a 25 Å radius hemisphere cut from a surface slab, with the inner 13 Å radius core allowed to relax, and 269 point charges around the outer shell to account for the charges of the missing slab atoms. Most calculations used a QM cluster consisting of 12 atoms, passivated by 13 Zn pseudopotentials and classical O anions. The MM potential used was previously developed, shown to be accurate for the structure, vibrational, elastic, dielectric and piezoelectric properties of bulk ZnO, and tested against QM calculations of small clusters and reconstructed polar surfaces. Energies of O defects were calculated, and the binding of electrons at the defect was found to be strongly favoured energetically. The default cluster size (12 atoms) was compared with a larger cluster (43 atoms) (see below), and the electron binding energies showed small but significant changes. The geometry of the defect sites, with and without the trapped electron, was also analysed. This system was also used as a basis for some tests of the method. For a

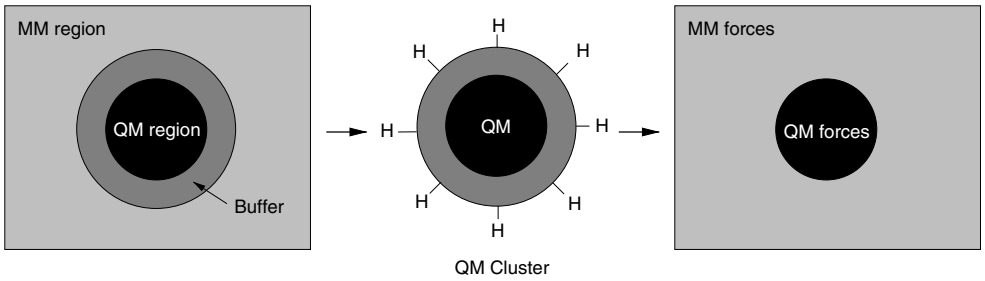

Figure 11. Schematic overview of the learn-on-the-fly scheme. A QM region is defined and then expanded to form a buffer region. The QM and buffer regions together form then a QM cluster, with the inaccurate QM forces on the buffer atoms being discarded. Finally MM and QM forces are assembled to give momentum-conserving forces for the entire system.

polar ZnO surface the external electrostatic potential applied to the QM region, including contributions from the inner and outer MM regions and a spherical charge shell, reproduces the Madelung potential in the QM region to within 10^{-4} V. For the application to electron trapping at the ZnO surface, the difference between using 12 atom and 43 atom QM clusters is 0.13–0.59 eV in binding energies of 2–4 eV, but no evidence of convergence w.r.t. the cluster size is shown. Anisotropic spin-dipole couplings (relevant for EPR experiments) vary by up to about 10% for the same two clusters.

The authors also modelled formate, an intermediate in the synthesis of methanol, bound to the ZnO catalyst polar O-terminated surface [30]. Similar size systems were used as for the ZnO surface O vacancy simulation. The simulations show that the formate adsorption leads to polarization of the molecule. The vibrational frequencies show reasonable agreement with experiment, but are actually independent of embedding.

Next, the coupling method was applied to the decomposition of N_2O, a pollutant, in Cu-exchanged zeolites. The system consisted of a 1100–1300 MM atom sphere and 5 or more atoms in the QM cluster. A benchmark cluster and embedded calculations for a small unit cell zeolite, chabazite, with a 36 atom basis, were compared with bulk plane-wave band-structure calculations. Most relaxed bond lengths in the cluster are improved by embedding (as compared with the isolated cluster calculation), but deviations of about 4% remain in the Cu–O bonds. However, the Si–O bonds at the boundary of the cluster adjacent to the Si–H passivation layer, which are 1.69 Å in the gas-phase cluster, increase to 1.70–1.72 Å in the embedded simulation, while the bulk calculation shows a length of 1.63 Å. The O adsorption energy differs from the bulk calculation by about 3.5 kcal mol^{-1}, which the authors describe as 'good agreement', but it is of the opposite sign (+1.9 kcal mol^{-1} versus −1.6 kcal mol^{-1}).

To determine the structure of the experimentally relevant zeolite, ZSM5 (288 atom basis) with adsorbed Cu and N_2O, the authors considered several possible sites, with unspecified QM cluster sizes, and relaxed all MM atoms within four hops of the Cu–O bond (45–68 relaxed MM atoms). They find results in good agreement with another embedded calculation [70], agreeing on the most stable Cu site with Cu–O bond lengths within 2%, although the (small) adsorption energy is twice as large in the QUASI work. No comparison with experiment is possible because experimental results are very uncertain, especially with respect to the location of the Cu atoms in the zeolite framework. The N_2O geometries and adsorption energies are discussed, although there are no comparisons with other calculations. Some decomposition reaction energies are also calculated, and these show significant differences from small bare cluster calculations that used a different DFT functional.

5.9. 'Learn-on-the-fly'

The 'learn-on-the-fly' (LOTF) scheme [20, 71] is a force-mixing method with a twist. As usual in force-mixing schemes, the QM forces are computed accurately by carving a cluster that includes a buffer region. The forces on all the buffer atoms are discarded and the remaining QM forces are merged with the empirical forces in the MM region—as illustrated in figure 11—making this an 'abrupt' force-mixing method. However, the resultant set of forces is not used directly to propagate the MD. Rather, a simple energy functional is constructed for the entire system, whose parameters (free to change from atom to atom) are fitted essentially continuously to best approximate the forces from the QM and MM calculations. The upshot of this is that the forces derived from the fitted energy functional sum to zero (in contrast to the merged QM and MM forces), so the dynamics conserves momentum. Furthermore, if the deviation of the QM and MM potential energy surfaces is not too large, the refitting procedure (and by extension the QM calculation which provides the input data for it) does not have to be carried out at every dynamical step. In the case of silicon, for example, a factor of ten in simulation speed can be gained this way without significant loss of accuracy, as illustrated in figure 12. Just as with other force-mixing schemes, there is evidently no conservation of total energy, due to the time dependence, through the refitting procedure, of the simple Hamiltonian from which the dynamical forces are derived. The form of the fitted energy functional has evolved considerably since the first introduction of this scheme. In [71] the empirical potential itself is used for this purpose, by refitting its parameters on a site-by-site basis in and near the QM region. Later it was recognized that in cases where the MM functional form is itself quite complicated, its refitting would be cumbersome and costly and is actually not necessary for the operation of the scheme. Instead, a simple short-range two-body potential $E_{\rm corr}$ is added to the unmodified MM potential and parameters of the former

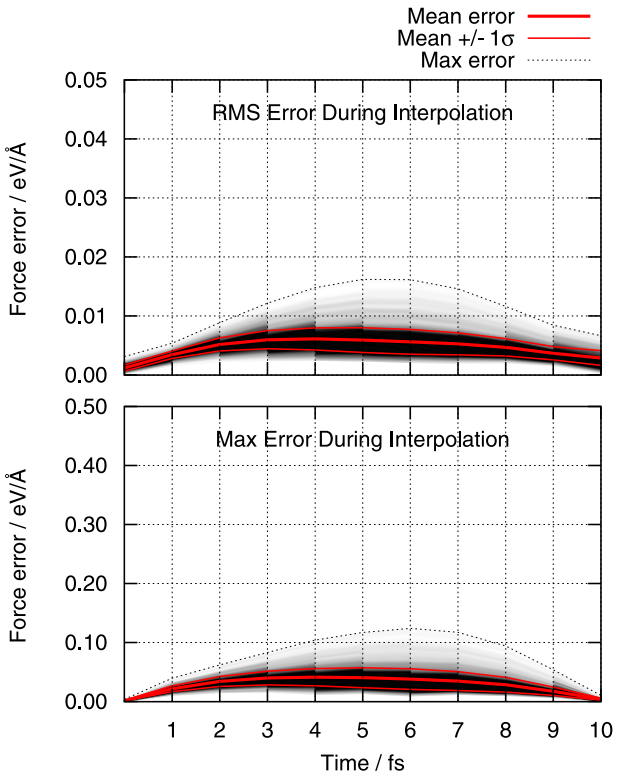

Figure 12. Force errors during a single LOTF cycle, averaged over many independent trajectories. Quantum mechanical calculations are only carried out at steps zero and ten, and the dynamics in between is interpolated. The test system is a 64 atom silicon bulk cube with a five atom QM region, at a temperature of 2000 K, with MD time step $\Delta t = 1$ ps. The grey-scale colouring shows the density of force errors as a function of time, with black corresponding to high and white to low densities. The mean and standard deviations of the distribution are indicated by the grey lines and the extreme values by the black dashed lines. The QM force model is tight binding, from [73].

are fitted to reproduce the force *difference* between the QM and MM models,

$$E_{\text{corr}} = \sum_{\langle i,j \rangle} V_2(\alpha_{ij}, r_{ij}), \tag{6}$$

$$\frac{\partial E_{\text{corr}}}{\partial x_i} \approx F_{\text{QM}} - F_{\text{MM}}, \tag{7}$$

where the indices i and j range over a set of atoms somewhat larger than the QM region and the parameters α_{ij} are optimized at each fitting step. While V_2 can be thought of as a corrective potential, the only data for its optimization at a particular instant in time are the set of QM and MM force differences for the current configuration, and it turns out that no accuracy is lost if V_2 is taken to be linear,

$$V_2(\alpha_{ij}, r_{ij}) = \alpha_{ij} r_{ij}.$$

Such a form lends itself to very fast optimization of the parameters α_{ij} using linear algebra. In the case where the QM calculations and the corresponding parameter fits are only carried out at every few dynamical steps, an alternative way of looking at this, quite apart from its hybrid aspect, is as a time-extrapolation scheme. The fundamental reason for its ability to 'beat' the largest possible Verlet time step is that the set of interatomic distances $\{r_{ij}\}$ are intrinsically very good coordinates for covalently bonded atomic systems, a fact that is widely appreciated [72].

The LOTF scheme has been extensively tested for numerical stability and accuracy in silicon. In [20] it was shown that the hybrid scheme, applying quantum mechanics (empirical tight binding) around a vacancy within a radius of 7.5 Å and the SW model elsewhere, correctly reproduces the diffusion coefficient of the defect with respect to the fully quantum mechanical calculation at several different temperatures. The buffer region was 7 Å thick, a value which was arrived at after convergence tests (explicitly shown in a later paper [33]). Several different parametrizations of the empirical tight-binding model were tried, which resulted in wide ranging diffusivities, the hybrid scheme in each case matching the quantum mechanical model that it was given. The tight-binding diffusivities not only differed from each other by more than an order of magnitude, but also in slope as a function of the inverse temperature, underscoring the severe limitations of tight-binding models for accurate simulation work. The glide of a partial dislocation was also simulated in a 4000+ atom unit cell, with the QM region around the kink defects on the dislocation line (which were forced to be present due to the skewed simulation cell and periodic boundary conditions). The stable configuration of the kink at high temperature was shown to be different using the hybrid scheme from the configuration at low temperature or that from a pure MM simulation (at any temperature), although no further conclusions were drawn from this.

A new quantitative test was presented in [33] for the case of the mechanical failure of a small rod of silicon containing a vacancy. The yield stress under the hybrid scheme with the QM region around the vacancy agreed well with the fully quantum mechanical result (and both were different from the pure MM value); however, similarly to the case of the CNT mentioned in section 5.6, the strain at failure was 20% lower. In this case, this discrepancy was traced to a surface rearrangement in the fully QM calculation that was not captured by the hybrid simulation because it occurred well away from the vacancy and hence outside the QM region. Further tests of the hybrid scheme included the correct surface reconstruction of the (1 0 0) surface, which the interatomic potential does not produce, and the qualitative, yet significant result that when using the hybrid scheme, a crack in Si propagates in a brittle manner with atomically flat clean crack surfaces.

This last result was the forerunner of the recently published study of crack path instabilities in the brittle fracture of silicon [7]. The system consisted of over 200 000 MM silicon atoms described using the SW potential and around 200 QM atoms, described using DFT as implemented in the CASTEP package [27]. The elliptical QM region was seeded at the crack tip and updated continuously using the hysteresis algorithm described above. The parameters for the updating algorithm were carefully tuned to allow the QM region to closely track the crack tip during fracture and at the same time not waste computing resources on regions

which have already cracked. An averaging time of 50 fs and inner and outer hysteretic ellipses of radii 8.0 Å and 9.3 Å, respectively, give a stable QM region which accurately follows the moving tip. Typically about 150 atoms were treated with QM at any given instant. QM calculations and corresponding optimizations for the corrective potential were performed after every 10 dynamical steps. It was discovered that the tips of atomically sharp cracks on the (1 1 1) surface propagating in the [1 1 $\bar{2}$] direction spontaneously reconstruct under load, leading to an instability for low speed cracks and macroscopically observable features of the crack surface. At higher speeds, the trajectory of the crack system in phase space guides it away from the reconstruction direction, and the crack path is smooth. The reconstruction leads, with some probability, to a step on the crack surface but always in the same direction. This fact allowed the atomic scale instability to be observed experimentally as micrometre scale ridges on the crack surface, always occurring along the same crystallographic orientation. Crack-tip reconstructions were suggested to be a general phenomenon occurring in covalent crystals under load, and preliminary results showing similar behaviour were presented for diamond and silicon carbide.

The cracks on the (1 1 0) surface also have instabilities which are well known, but at higher speeds in this case, and the simulations suggested an atomistic explanation. In this orientation there is a near degeneracy between the breaking of bonds above and below the crack plane. At low speeds the weakest bonds always break first, and the crack propagates in a straight path along the (0 0 1) direction. At higher speeds the energy released when the bonds break enables the crack to break the slightly stronger bonds, thereby diverting it onto one of two oblique [1 $\bar{1}$ 1] planes. A small amount of in-plane shear breaks the symmetry between the two [1 $\bar{1}$ 1] planes, causing the crack to systematically diverge from its original (1 1 0) orientation.

5.10. Embedding with lattice GFs

Woodward *et al* embed a QM region in an effective infinite atomistic region for a particular application, to get the relaxed structure of an isolated dislocation which is periodic along the dislocation line [66, 74–77]. They do this by applying flexible boundary conditions determined by the lattice GF [78] (the displacements of neighbouring atoms necessary to produce a particular force on an atom). The atomistic system consists of a cylinder surrounding the dislocation core, partitioned into three regions with increasing outer radii: a QM region near the dislocation, a coupling region and a buffer region, as schematically illustrated in figure 13. Positions are initialized to the analytic solution corresponding to the anisotropic continuum displacement field. Finding the relaxed structure proceeds by alternating steps. First the coupling and buffer atoms are constrained and the QM atoms relaxed according to the QM forces, computed in a system that includes all the atoms. Then all the atoms are moved according to the displacements determined by the coupling atom *QM* forces and the lattice Green's function (lGF). These two steps are repeated until forces on the QM and coupling atoms are minimized to some specified accuracy.

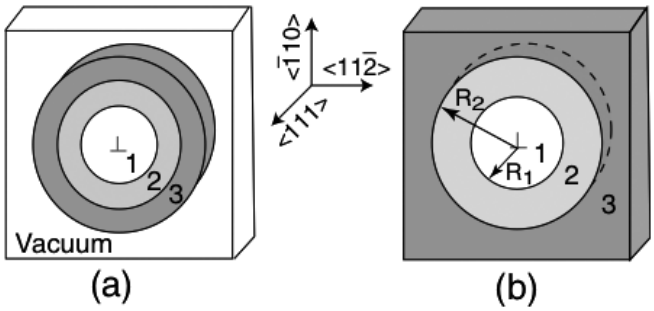

Figure 13. Schematic overview of the lattice GF embedding method, showing the two types of simulation cells used to model dislocations. In (*a*) a vacuum region is used to terminate the dislocation and in (*b*) a domain boundary is used. Reprinted with permission from [66]. Copyright 2008 by the American Physical Society.

While this method is very different from other embedding methods discussed here, similar considerations apply because the method depends on two things: an accurate lGF (analogous to QM–MM property matching) and accurate QM forces in the QM and coupling regions (i.e. minimizing edge effects in the QM region). The lGF is closely related to the force-constant matrix and can be computed for the QM energy from finite differences or a linear response calculation in a periodic unit cell. Special efforts must be made to correctly account for the fact that practical unit cells are smaller than the range of force constants in DFT simulated metals, and information about the elastic constants can also be used [79]. The effects of the boundary on forces in the QM and coupling regions have been discussed in some detail [66, 74–77]. Both free boundaries and square systems in periodic supercells (with defects caused by the incompatibility between the dislocation and the periodicity) have been used in the past [76], but the latest work uses free boundaries [66].

The method has been applied to dislocations in body-centred cubic Mo and Ta, γ-TiAl and face-centred cubic Al [66,75–77], computing the relaxed core structure and Peierls stress, i.e. the stress needed to glide the dislocation at $T = 0$ K. It has been difficult to test the complete method, because the DFT calculation cannot be converged with respect to the unit cell size. The underlying lGF method has been tested for interatomic potentials [74], but the DFT based implementations have focused on testing the convergence of the QM forces with respect to the region size. For Mo and Ta screw dislocations, atom positions near the core were converged to within 0.01 Å with respect to QM and the coupling region radius. For Al the authors indicate that a 24 Å cylinder outer radius is sufficient and that forces on atoms in the outermost 10 Å shell are affected by the free surface (to an unspecified accuracy).

The method was used to simulate the structure of the screw dislocation in Mo and Ta and found that it was qualitatively different from previous interatomic potential calculations for Mo [75]. Comparison of the glide stress in Mo for soft and hard directions showed reasonable agreement with experiment. For γ-TiAl $\frac{1}{2}$[1 1 0] screw dislocation simulations found a non-planar dislocation core, with double the yield stress found in experiment, and support for a model for the

temperature dependence of the yield stress. For screw and edge dislocations in Al the method predicted dissociation into partials with separations in good agreement with experiment, in contrast to previous interatomic potential based simulations that predict widely varying and generally too high partial separations.

6. Discussion

The ten methods we reviewed above use a range of formulations to couple the QM and MM regions, but most can be classified as force mixing or ONIOM-like energy-mixing. Despite the growing number of methods and applications, few of the publications have included quantitative tests of the accuracy of the method and its convergence with respect to possible parameters. In order to compare the accuracy of the coupling methods, we carried out a test using our own representative implementations of some of the approaches. The behaviour of a simulated system depends on the potential energy surface (or some effective potential energy surface in the case of force mixing). As a generic probe of this potential energy surface, we use its gradient, i.e. the atomic forces, at a particular localized deformation. Since one can compute energy differences (between different minima or between a minimum and an adjacent saddle point) by integrating the force difference over a trajectory, the accuracy of the force in a particular method is indicative of the accuracy of the method at simulating the dynamical evolution of the system. We actually measure average forces at a finite temperature, because the corresponding *free energy surface* controls the dynamics and observables at finite temperatures and this also enables us to sample the errors of the various schemes in a large number of representative configurations.

Specifically, we simulate a 1000 atom cube of crystalline silicon using the libAtoms code [80], with one bond length expanded from the equilibrium value of 2.35 to 2.65 Å. We measure the average force needed to keep that constraint satisfied during a 100 ps MD run at $T = 300$ K. The reference value, which a good coupling method should reproduce, is computed using a fully QM simulation evaluating the total energy using Bowler's TB parametrization [81] and exact diagonalization. We compare the full QM simulation with a number of coupling methods: LOTF, three variants of force-mixing and ONIOM, all using the same QM method as the full QM simulation coupled to the SW potential using the published parameters [17]. The LOTF and ONIOM methods are described above. Note that for the original ONIOM there is no possibility of using a buffer region. This is not the case for Ogata's modification, but we do not at present have an implementation of that method, so we were unable to test it. All of the force-mixing variants begin by computing the forces in the entire system with the interatomic potential. The QM calculation includes the strained bond, a core defined by a number of hops from the two strained bond atoms, a buffer region (also defined in terms of bond hops) surrounding the core and a layer of passivation H atoms. Abrupt force mixing simply uses QM forces in the QM region and MM forces elsewhere. Smooth force mixing uses a weighted sum of QM and MM forces in the buffer region, with the weights linearly interpolating from 1 to 0 as a function of the number of *hops* from the core for the hop-ramp and as a function of the *distance* from the core for the distance-ramp. Momentum-conserving (CM) force mixing is similar to abrupt force mixing, but a small correction force is added to every QM force to conserve the total momentum of the system at every time step.

The average values of the deviation of the constraint force from the full QM simulation are plotted in figure 14, as a function of the QM region size and the buffer size. A number of trends are clear in the plots. The ONIOM method, the only one in this test that properly conserves energy, has the smallest error when buffer regions are not employed, but this error is still substantial and does not improve as a function of the one systematic parameter available, the size of the QM region. For the same overall computational effort that is implied by a QM region size of 3 hops, a more accurate average force can be computed by employing a QM region of 1 hop and a buffer region of 2 hops. The LOTF method gives smaller errors that are systematically reduced by increasing the buffer region, but it is not noticeably better than any of the much simpler force-mixing variants. At a given computational cost (i.e. the total number of atoms in the QM calculation, which includes the QM and buffer regions), abrupt force mixing is better than smooth force mixing. We speculate that this is because atoms near the QM cluster boundary have forces very different from the correct force on the analogous atoms in the complete system. Apparently the disturbance caused by mixing those incorrect forces with the MM forces is larger than any benefit from a smooth transition from QM to MM forces. Finally, the imposition of momentum conservation by adding a small force to every QM atom does not make a significant change to the constraint force.

Another measure of coupling accuracy relevant to the force-mixing variants is how well the energy is conserved. Since the total energy is not defined, we measure the average temperature (i.e. kinetic energy) in the QM region in a Langevin-thermostated [82] MD run without constraints. If energy is conserved, the thermostat should be equally effective in all of the simulated system, and the QM region temperature should be the same as the bulk temperature and the thermostat target temperature. If energy conservation is violated, for example, near the QM/MM region boundaries, the QM region could be hotter or colder than the thermostat temperature. We plot the QM region temperature for the LOTF method and for force-mixing using abrupt, hop-ramp and momentum conservation variants in figure 15. The temperature deviation is positive and largest in magnitude for hop-ramp force mixing. With abrupt force mixing and LOTF the QM region temperature deviation is much smaller and negative, but the best results are for abrupt force mixing with imposed momentum conservation (note that momentum conservation is guaranteed by construction for LOTF).

We conclude with some thoughts on possible future developments of hybrid schemes for solid-state systems. It is clear from the above test that the presently available force-mixing schemes have superior accuracy over the widely

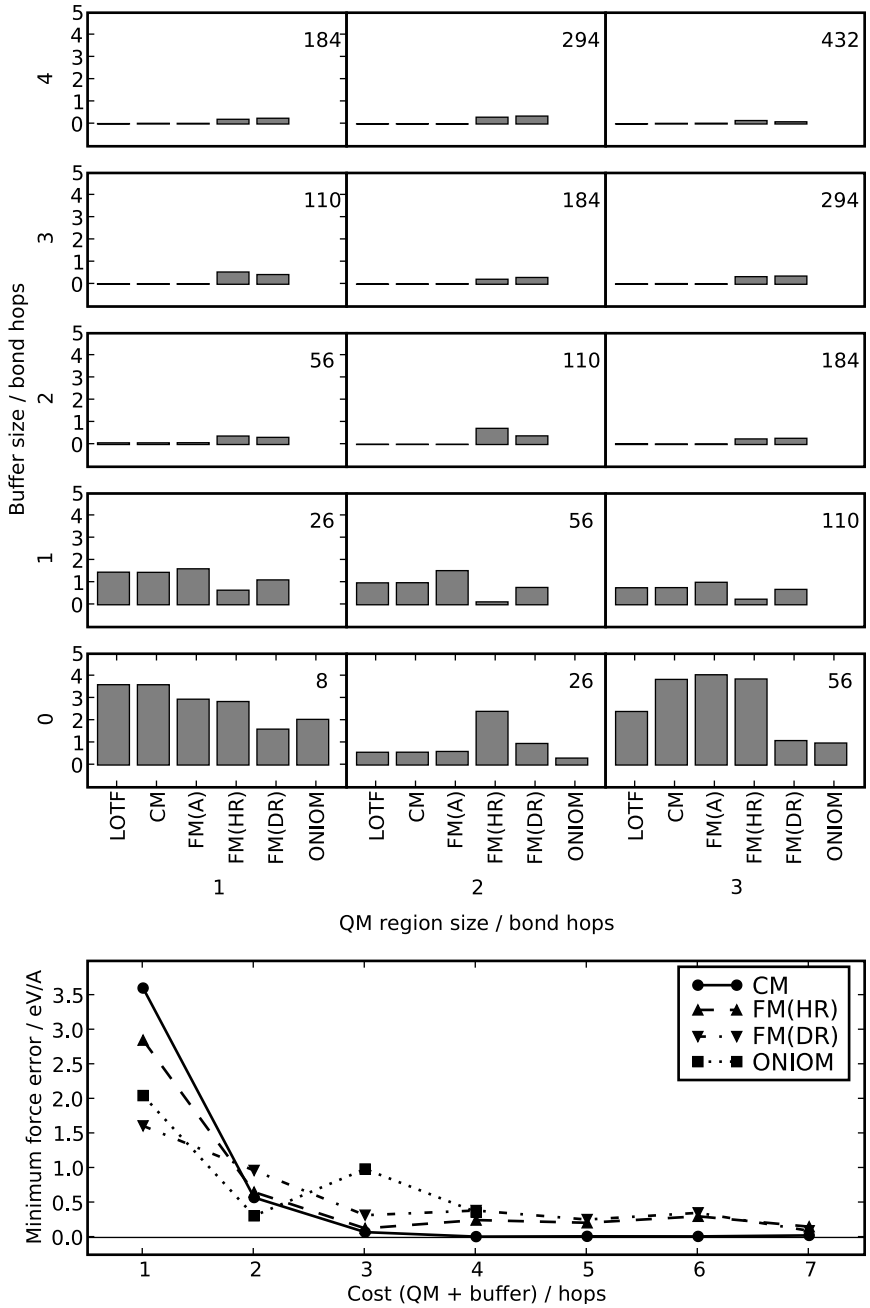

Figure 14. Deviation between fully QM simulation and various coupling methods in the average force required to keep a bond in bulk silicon constrained to a length of 2.65 Å during a $T = 300$ K MD simulation. QM energies and forces are computed using Bowler's TB parametrization, and MM energies and forces are computed using the SW potential. The coupling methods are learn-on-the-fly (LOTF), abrupt force mixing with conservation of momentum (CM), abrupt force-mixing (FM(A)), force mixing with a hop-ramp (FM(HR)), force mixing with a distance-ramp (FM(DR)) and ONIOM. Upper panel: deviation is plotted for each method as a function of the QM region core size and the QM buffer region size, defined by the number of bond hops. The number in each panel indicates the total number of Si atoms in the QM calculation (passivation H atoms are not included). Lower panel: force deviation as a function of the total QM size for a subset of these methods (the LOTF and FM(A) methods are not shown as they are almost identical to the CM method). For each method the point shown corresponds to the choice of buffer size which gives the smallest deviation from the fully QM force on the constrained bond, for ONIOM the buffer size is zero by definition.

applicable ONIOM-type energy-mixing scheme. However, it seems unlikely that energy conservation could easily be restored into force-mixing schemes, and therefore disturbing questions linger in evaluating any application in connection with local heating and how much that could affect the final results. On the other hand, there is scope for developing energy-mixing schemes that use buffer regions, and hence they could achieve similarly high accuracy. In order to exclude the energy of the buffer region (which is obviously wrong), a generally applicable method must be found to partition the quantum mechanical total energy spatially in a physically meaningful way.

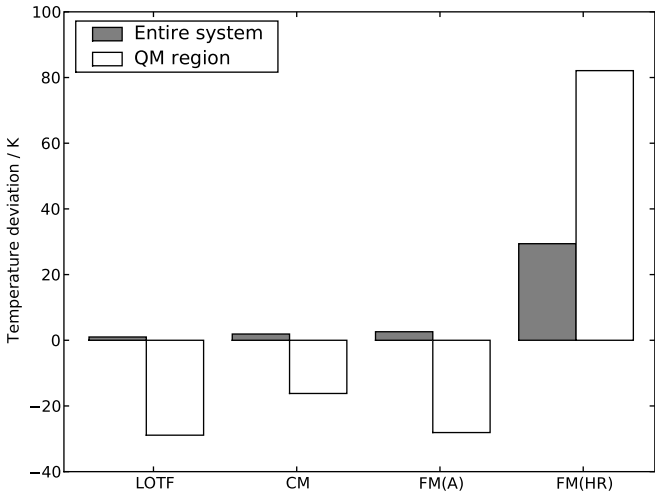

Figure 15. Temperature deviations from the Langevin thermostat target of 300 K in the entire system and in the QM core region for various force-mixing based coupling methods. The system is a 1000 atom crystal of bulk silicon, with a QM core of 26 atoms and a buffer region of three bond hops, leading to QM clusters containing 184 silicon atoms plus passivation H atoms. Temperatures were obtained by averaging over 100 ps MD runs with a thermostat time constant of 1 ps. The force-mixing methods are as described in figure 14.

Acknowledgments

The authors would like to acknowledge helpful discussions with Steven Winfield who suggested the momentum-conserving abrupt force-mixing method. Computing resources were provided by the HPCS at the University of Cambridge. NB would like to thank the NRL and the ONR for providing funding and the US DOD HPCMPO for computing resources. GC and JRK would like to acknowledge funding from the EPSRC under Grant no EP/C52392X/1.

References

[1] Probert M and Payne M 2003 Improving the convergence of defect calculations in supercells: an ab initio study of the neutral silicon vacancy *Phys. Rev.* B **67** 075204

[2] Tadmor E B, Ortiz M and Phillips R 1996 Quasicontinuum analysis of defects in solids *Phil. Mag.* A **73** 1529–63

[3] Miller R, Ortiz M, Phillips R, Shenoy V and Tadmor E B 1998 Quasicontinuum models of fracture and plasticity *Eng. Fract. Mech.* **61** 427–44

[4] Shenoy V B, Miller R, Tadmor E B, Rodney D, Phillips R and Ortiz M 1999 An adaptive finite element approach to atomic-scale mechanics—the quasicontinuum method *J. Mech. Phys. Solids* **47** 611–42

[5] Rappe A K and Goddard W A 1991 Charge equilibration for molecular dynamics simulations *J. Phys. Chem.* **95** 3358–63

[6] Jorgensen W L 2007 Special issue on polarization *J. Chem. Theory Comput.* **3** 1877

[7] Kermode J, Albaret T, Sherman D, Bernstein N, Gumbsch P, Payne M C, Csanyi G and De Vita A 2008 Low speed fracture instabilities in a brittle crystal *Nature* **455** 1224–7

[8] Frenkel D and Smit B 2001 *Understanding Molecular Simulation* (*Computational Science Series* vol 1) (New York: Academic)

[9] Broughton J Q, Abraham F F, Bernstein N and Kaxiras E 1999 Concurrent coupling of length scales: methodology and application *Phys. Rev.* B **60** 2391–403

[10] Bernstein N and Hess D W 2003 Lattice trapping barriers to brittle fracture *Phys. Rev. Lett.* **91** 025501

[11] Khare R, Mielke S L, Paci J T, Zhang S, Ballarini R, Schatz G C and Belytschko T 2007 Coupled quantum mechanical/molecular mechanical modeling of the fracture of defective carbon nanotubes and graphene sheets *Phys. Rev.* B **75** 075412

[12] Choly N, Lu G, Weinan E and Kaxiras E 2005 Multiscale simulations in simple metals: a density-functional-based methodology *Phys. Rev.* B **71** 094101

[13] Lu G, Tadmor E and Kaxiras E 2006 From electrons to finite elements: a concurrent multiscale approach for metals *Phys. Rev.* B **73** 024108

[14] Weinan E 2006 Uniform accuracy of the quasicontinuum method *Phys. Rev.* B **74** 12

[15] Shenoy V B, Miller R, Tadmor E B, Phillips R and Ortiz M 1998 Quasicontinuum models of interfacial structure and deformation *Phys. Rev. Lett.* **80** 742–5

[16] Srivastava G P 1990 *The Physics of Phonons* (Boca Raton, FL: CRC Press)

[17] Stillinger F H and Weber T A 1985 Computer simulation of local order in condensed phases of silicon *Phys. Rev.* B **31** 5262

[18] Frauenheim Th, Seifert G, Elsterner M, Hajnal Z, Jungnickel G, Porezag D, Suhai S and Scholz R 2000 A self-consistent charge density-functional based tight-binding method for predictive materials simulations in physics, chemistry and biology *Phys. Status Solidi* b **217** 41

[19] Spence J C H, Huang Y M and Sankey O 1993 Lattice trapping and surface reconstruction for silicon cleavage on (1 1 1). Ab-initio quantum molecular dynamics calculations *Acta Metall. Mater.* **41** 2815–24

[20] Csanyi G, Albaret T, Payne M C and De Vita A 2004 'Learn on the fly': a hybrid classical and quantum-mechanical molecular dynamics simulation *Phys. Rev. Lett.* **93** 175503

[21] Goedecker S 1998 Decay properties of the finite-temperature density matrix in metals *Phys. Rev.* B **58** 3501

[22] Kohn W 1996 Density functional and density matrix method scaling linearly with the number of atoms *Phys. Rev. Lett.* **76** 3168

[23] Ismail-Beigi S and Arias T 1999 Locality of the density matrix in metals, semiconductors, and insulators *Phys. Rev. Lett.* **82** 2127

[24] Ogata S, Lidorikis E, Shimojo F, Nakano A, Vashishta P and Kalia R K 2001 Hybrid finite-element/molecular-dynamics/electronic-density-functional approach to materials simulations on parallel computers *Comput. Phys. Commun.* **138** 143–54

[25] Ogata S 2005 Buffered-cluster method for hybridization of density-functional theory and classical molecular dynamics: application to stress-dependent reaction of H_2O on nanostructured Si *Phys. Rev.* B **72** 045348

[26] Laio A, VandeVondele J and Rothlisberger U 2002 A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations *J. Chem. Phys.* **116** 6941–7

[27] Segall M D, Lindan P J D, Probert M J, Pickard C J, Hasnip P J, Clark S J and Payne M C 2002 First-principles simulation: ideas, illustrations and the CASTEP code *J. Phys.: Condens. Matter* **14** 2717–43

[28] Bowler D and Gillan M 2002 An embedding scheme based on quantum linear-scaling methods *Chem. Phys. Lett.* **355** 306–10

[29] Sokol A A, Bromley S T, French S A, Catlow C R A and Sherwood P 2004 Hybrid QM/MM embedding approach for the treatment of localized surface states in ionic materials *Int. J. Quantum Chem.* **99** 695–712

[30] Sherwood P *et al* 2003 QUASI: a general purpose implementation of the QM/MM approach and its application to problems in catalysis *J. Mol. Struct.: THEOCHEM* **632** 1–28

[31] Svensson M, Humbel S, Froese R D J, Matsubara T, Sieber S and Morokuma K 1996 ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. a test for Diels–Alder reactions and Pt(P(*t*-Bu)$_3$)$_2$ + H_2 oxidative addition *J. Phys. Chem.* **100** 19357–63

[32] Bernstein N 2001 Linear scaling nonorthogonal tight-binding molecular dynamics for nonperiodic systems *Europhys. Lett.* **55** 52–8

[33] Csanyi G, Albaret T, Moras G, Payne M C and De Vita A 2005 Multiscale hybrid simulation methods for material systems *J. Phys.: Condens. Matter* **17** R691–703

[34] Warshel A and Levitt M 1976 Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme *J. Mol. Biol.* **103** 227–49

[35] Perez R and Gumbsch P 2000 Directional anisotropy in the cleavage fracture of silicon *Phys. Rev. Lett.* **84** 5347–50

[36] Rice J R 1992 Dislocation nucleation from a crack tip: an analysis based on the Peierls concept *J. Mech. Phys. Solids* **40** 239–71

[37] Kaxiras E and Duesbery M S 1993 Free energies of generalized stacking faults in Si and implications for the brittle–ductile transition *Phys. Rev. Lett.* **70** 3752–5

[38] Hauch J A, Holland D, Marder M P and Swinney H L 1999 Dynamic fracture in single crystal silicon *Phys. Rev. Lett.* **82** 3823–6

[39] Ogata S, Shimojo F, Kalia R K, Nakano A and Vashishta P 2002 Hybrid quantum mechanical/molecular dynamics simulation on parallel computers: density functional theory on real-space multigrids *Comput. Phys. Commun.* **149** 30–8

[40] Ogata S, Shimojo F, Kalia R K, Nakano A and Vashishta P 2004 Environmental effects of H_2O on fracture initiation in silicon: a hybrid electronic-density-functional/molecular-dynamics study *J. Appl. Phys.* **95** 5316–23

[41] Ciacchi L and Payne M 2005 First-principles molecular-dynamics study of native oxide growth on Si(0 0 1) *Phys. Rev. Lett.* **95** 196101

[42] Belkada R, Igarashi T and Ogata S 2004 Effects of H_2O on Si fracture: a hybrid quantum-classical simulation *Comput. Mater. Sci.* **30** 195–201

[43] Cho J-H, Kim K, Lee S-H and Kang M-H 2000 Dissociative adsorption of water on the Si(0 0 1) surface: a first-principles study *Phys. Rev.* B **61** 4503

[44] Akagi K and Tsukada M 1999 Theoretical study of the hydrogen relay dissociation of water molecules on Si(0 0 1) surfaces *Surf. Sci.* **438** 9–17

[45] Abraham F, Bernstein N, Broughton J and Hess D 2000 Dynamic fracture of silicon: concurrent simulation of quantum electrons, classical atoms and the continuum solid *MRS Bull.* **25** 27

[46] Bazant M Z, Kaxiras E and Justo J F 1997 Environment-dependent interatomic potential for bulk silicon *Phys. Rev.* B **56** 8542–52

[47] Bernstein N and Kaxiras E 1997 *Phys. Rev.* B **56** 10488

[48] Nieminen J A and Paavilainen S 1999 Bridging the gap over size scales: a Greens-function method to combine tight-binding and semiempirical force calculations *Phys. Rev.* B **60** 2921–9

[49] Sutton A P and Chen J 1990 Long-range Finnis Sinclair potentials *Phil. Mag. Lett.* **61** 139–46

[50] Rafiii-Tabar H and Sutton A P 1991 Long-range Finnis-Sinclair potentials for fcc metallic alloys *Phil. Mag. Lett.* **63** 217–24

[51] Lannoo M and Friedel P 1991 *Atomic and Electronic Structure of Surfaces* (Berlin: Springer)

[52] Fridel J 1969 *The Physics of Metals* ed J M Ziman (New York: Cambridge University Press)

[53] Feynman R P 1939 Forces in molecules *Phys. Rev.* **56** 340

[54] Hellmann H 1937 *Einfurung in die Quantenchemie* pp 61, 285 (Leipzig: Deuticke)

[55] Tanaka H, Yoshinobu J and Kawai M 1995 Oxygen-induced reconstruction of the Pd(1 1 0) surface: an STM study *Surf. Sci.* **327** L505–9

[56] Pussi K, Hirsimäki M, Valden M and Lindroos M 2004 Multilayer relaxation of Pd{3 2 0} surface by quantitative LEED revisited *Surf. Sci.* **566–568** 24–8

[57] Schmidt M W *et al* 1993 *J. Comput. Chem.* **14** 1347–63

[58] Sanchez-Portal D, Ordejón P, Artacho E and Soler J M 1997 Density-functional method for very large systems with LCAO basis sets *Int. J. Quantum Chem.* **65** 453

[59] Moran B and Shih C 1987 Crack tip and associated domain integrals from momentum and energy balance *Eng. Fract. Mech.* **27** 615–42

[60] Chen Z, Nagase S, Hirsch A, Haddon R C, Thiel W and Von Ragué Schleyer P 2004 Side-wall opening of single-walled carbon nanotubes (SWCNTs) by chemical modification: a critical theoretical study *Angew. Chem. Int. Ed. Engl.* **43** 1552–4

[61] Yim W L and Liu Z F 2004 A reexamination of the chemisorption and desorption of ozone on the exterior of a (5,5) single-walled carbon nanotube *Chem. Phys. Lett.* **398** 297–303

[62] Petsalakis I, Pagona G, Theodorakopoulos G, Tagmatarchis N, Yudasaka M and Iijima S 2006 Unbalanced strain-directed functionalization of carbon nanohorns: a theoretical investigation based on complementary methods *Chem. Phys. Lett.* **429** 194–8

[63] Khare R, Mielke S, Schatz G and Belytschko T 2008 Multiscale coupling schemes spanning the quantum mechanical atomistic forcefield, and continuum regimes *Comput. Methods Appl. Mech. Eng.* **197** 3190–202

[64] Daw M S and Baskes M I 1984 Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals *Phys. Rev.* B **29** 6443–53

[65] Ercolessi F and Adams J B 1994 Interatomic potentials from first-principles calculations: the force-matching method *Europhys. Lett.* **26** 583–8

[66] Woodward C, Trinkle D R, Hector L G Jr and Olmsted D L 2008 Prediction of dislocation cores in aluminum from density functional theory *Phys. Rev. Lett.* **100** 045507

[67] Mills M and Stadelmann P 1989 A study of the structure of Lomer and 60° dislocations in aluminium using high-resolution transmission electron microscopy *Phil. Mag.* A **60** 355–84

[68] Liu Y, Lu G, Chen Z and Kioussis N 2007 An improved QM/MM approach for metals *Modelling Simul. Mater. Sci. Eng.* **15** 275–84

[69] Chemshell, a Computational Chemistry Shell, see http://www.chemshell.org

[70] Sierka M, Nachtigall P and Sauer J 1999 *Phys. Chem. Chem. Phys.* **1** 2019

[71] De Vita A and Car R 1998 *Mater. Res. Soc. Symp. Proc.* **491** 473

[72] Pulay P and Fogarasi G 1992 Geometry optimization in redundant internal coordinates *J. Chem. Phys.* **96** 2856–60

[73] Kwon I, Biswas R, Wang C Z, Ho K M and Soukoulis C M 1994 Transferable tight-binding models for silicon *Phys. Rev.* B **49** 7242

[74] Rao S, Hernandez C, Simmons J, Parthasarathy T and Woodward C 1998 Green's function boundary conditions in two-dimensional and three-dimensional atomistic simulations of dislocations *Phil. Mag.* A **77** 231–56
[75] Woodward C and Rao S 2001 Ab-initio simulation of isolated screw dislocations in bcc Mo and Ta *Phil. Mag.* A **81** 1305–16
[76] Woodward C and Rao S 2002 Flexible ab initio boundary conditions: simulating isolated dislocations in bcc Mo and Ta *Phys. Rev. Lett.* **88** 216402
[77] Cawkwell M, Nguyen-Manh D, Woodward C, Pettifor D and Vitek V 2005 Origin of brittle cleavage in iridium *Science* **309** 1059–62
[78] Trinkle D 2008 Lattice green function for extended defect calculations: computation and error estimation with long-range forces *Phys. Rev.* B **78** 014110
[79] Trinkle D R 2007 Lattice and elastic constants of titanium-niobium monoborides containing aluminum and vanadium *Scr. Mater.* **56** 273–6
[80] Csanyi G, Winfield S, Kermode J, Comisso A, De Vita A, Bernstein N and Payne M C 2007 Expressive programming for computational physics in Fortran 95+ *IoP Comput. Phys. Newsletter* Spring 2007
[81] Bowler D R, Fearn M, Goringe C M, Horsfield A P and Pettifor D G 1998 Hydrogen diffusion on Si(0 0 1) studied with the local density approximation and tight binding *J. Phys.: Condens. Matter* **10** 3719
[82] Adelman S A and Doll J D 1976 Generalized Langevin equation approach for atom/solid-surface scattering: general formulation for classical scattering off harmonic solids *J. Chem. Phys.* **64** 2375–88