

Multiscale hybrid simulation methods for material systems

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2005 J. Phys.: Condens. Matter 17 R691

(<http://iopscience.iop.org/0953-8984/17/27/R02>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 129.252.86.83

The article was downloaded on 28/05/2010 at 05:13

Please note that [terms and conditions apply](#).

TOPICAL REVIEW

Multiscale hybrid simulation methods for material systems

Gabor Csányi¹, T Albaret², G Moras³, M C Payne¹ and A De Vita^{3,4}¹ Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK² Laboratoire de Physique de la Matière Condensée et Nanostructures, Bâtiment Leon Brillouin, Campus de la Doua, Université Claude Bernard Lyon 1, 43 Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France³ Physics Department, King's College London, Strand, London WC2R 2LS, UK⁴ INFM-DEMOCRITOS National Simulation Centre and Centre of Excellence for Nanostructured Materials (CENMAT), University of Trieste, Italy

Received 21 January 2005

Published 24 June 2005

Online at stacks.iop.org/JPhysCM/17/R691**Abstract**

We review recent progress in the field of multiscale hybrid computer simulations of materials, and present an overview of a novel scheme that links arbitrary atomistic simulation techniques together in a truly seamless manner. Rather than constructing a new hybrid Hamiltonian that combines different models, we use a unique short range classical potential and continuously tune its parameters to reproduce the atomic trajectories at the prescribed level of accuracy throughout the system.

Contents

1. Introduction	691
2. Hybrid schemes	692
3. New scheme	694
4. Validation	699
5. Brittle failure	700
6. Conclusion	702
Acknowledgments	703
References	703

1. Introduction

Natural phenomena occur on a variety of length scales, which not only roughly define scientific disciplines (physics and chemistry for the very small, biology for intermediate, geology for very long times and large length scales, and cosmology for the largest), but also delineate subfields within a given discipline, due to the very different experimental methods and theoretical models

that are applicable at each scale. These form a multiscale hierarchy, in which parameters of larger scale models are measured or calculated on a smaller scale [1].

Yet in a large class of problems, the length scales cannot be separated in this way; the coupling between them is strong and ‘bidirectional’. This often happens when the microscopic phenomena are driven by some macroscopic force. Stress induced defect processes in solids are a good example, and brittle fracture is the prototypical problem. In molecular biology, catalysis is often controlled by large scale motion of macromolecules. If we wish to simulate these processes, we have to do it on more than one length scale *simultaneously*. Crucially, one cannot use the smallest scale model to simulate the entire system, because it would be too expensive and at the same time hugely wasteful. A typical quantum mechanical molecular dynamics (MD) simulation can deal with hundreds of atoms, while the minimum system size that can capture the larger scale aspects can reach into hundreds of thousands of atoms. In the past decade there has been a growing effort to devise so-called *hybrid* or *embedded* dynamical simulation techniques that seamlessly integrate a wide variety of different models, ranging from first-principles methods to finite elements techniques, into a concurrent simulation.

In some cases, the processes to be represented on the larger scale are simple. A trivial case, for example, is that of a surface calculation, where one takes a thin slab of crystal, and keeps the atoms in the lowest layers fixed in their bulk positions. While, in some sense, this can be construed as an embedding scheme, it is more interesting when the large scale effect, while still simple, is not static but dynamic, e.g. maintaining isotropic pressure in the simulation of a small particle by putting it in a bath of soft spheres [2]. In most problems, however, the large scale behaviour is sufficiently complex that it requires a physically realistic description.

Pioneering work in this field was by Kohlhoff *et al* [3] and also by Tadmor *et al* who developed the *quasicontinuum* method [4], which successfully linked classical atomistic and continuum elasticity models. Here we concentrate on a lower level linkage: that of quantum mechanical to classical modelling. After mentioning general issues related to hybrid atomistic modelling and previous approaches, we discuss our scheme which was proposed in [5]. Finally, we present several validation tests and an example application to the brittle fracture in silicon.

2. Hybrid schemes

The objective is the following: given a large atomic system, some (perhaps small) regions of it need to be simulated with a quantum mechanical, and the rest with a classical empirical atomistic model. The major issues that have to be addressed are as follows.

- **Handshaking.** The two models need to interact at the border that separates the two regions. How do each of the models react to having an artificial surface where we switch to using the other model?
- **Selection.** How do we select which regions are to be treated by which model? In some cases, this might be straightforward, e.g. if we are investigating an impurity or a stationary defect. But in general, the region of interest where inherently quantum mechanical processes take place can appear, disappear and move, and we need to be able to track these changes.
- **Validation.** Given any particular hybrid scheme, it has to be validated against fully accurate quantum mechanical calculations. A successful scheme will be insensitive to the precise location of the border between the two models, and it should be possible to converge all observables by increasing the size of the quantum region.

Critically, the atomistic processes, as described by our quantum mechanical model, must be *local*. It is well known that in the vast majority of cases, quantum mechanics is local in

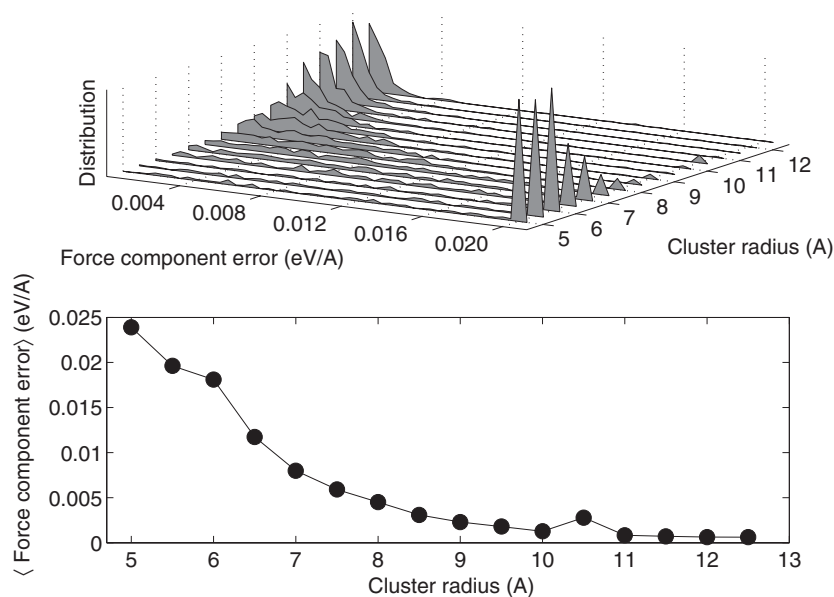


Figure 1. (a) Distribution of force component errors when using a finite cluster centred on the atom to compute the force, plotted for a range of cluster sizes, showing that for larger clusters the observed errors get progressively smaller. The peaks on the right represent the integrated distribution of errors larger than $0.02 \text{ eV } \text{Å}^{-1}$. The system is a silicon self-interstitial in a 512-atom cubic cell with periodic boundary conditions, equilibrated at 1400 K. Only atoms within 8 Å of the interstitial were considered for the statistics. (b) Mean of the absolute error in the force components for various cluster sizes. The peak at 10.5 is due to the periodicity of the underlying system. All clusters are terminated by hydrogen atoms, the force model is tight binding from [6].

the sense that the density matrix $\rho(\mathbf{r}, \mathbf{r}')$ is sparse; its elements decay fast as the separation between \mathbf{r} and \mathbf{r}' increases (this kind of locality we call *weak locality*). This is very useful for creating linear scaling electronic structure algorithms, but is not enough for a hybrid approach to be valid. What is needed is *strong locality*, which we define by

$$\frac{\partial^n}{\partial \mathbf{x}_j^n} \frac{\partial E_{\text{total}}}{\partial \mathbf{x}_i} \rightarrow 0 \quad \text{as} \quad |\mathbf{x}_i - \mathbf{x}_j| \rightarrow \infty \quad \forall n, i \neq j \quad (1)$$

where E_{total} is the total energy, and \mathbf{x}_i and \mathbf{x}_j are the atomic positions of atoms i and j . This guarantees that, if the quantum region is large enough, the trajectories that are important at the centre of the quantum region are not affected by the fact that far away, the system is treated classically (and hence the microscopic trajectories will not be very accurate) and also that the trajectories in the quantum region can be computed with high accuracy by only considering a finite neighbourhood. The length scale of the decay in equation (1) thus gives a guide to how large the quantum region needs to be. Most quantum mechanical systems either obey strong locality, or at least the parts of the Hamiltonian that do not, such as those that give rise to long range Coulomb forces and van der Waals forces, can be very effectively treated purely classically, outside the hybrid framework. By way of an example, figure 1 shows the distribution of errors in computing the force components on an atom using a finite cluster, as compared with using periodic boundary conditions, using a tight binding model [6].

There is a large body of work in the chemistry literature, collectively referred to as quantum mechanics/molecular mechanics (QM/MM), which tries to solve this problem in the context of biomolecular simulations (see [7] for a review). While that is not the subject here, we

note the fundamental difficulty of these techniques. When a quantum mechanical calculation of a subsystem is incorporated into the total Hamiltonian, the artificially broken bonds on the surface of the subsystem are generally terminated with hydrogen atoms or some kind of pseudo-atom that is supposed to mimic the electronic effect of the region outside the subsystem that has been removed. However, typical length scales of strong locality for the electronic part of the energy are of the order of a nanometre, which means that no matter how complicated the termination strategy is, as long as it involves just replacing a broken bond, the resulting forces near the edge of the quantum mechanically treated region will be incorrect. It is interesting to note that the success of higher level hybrid schemes mentioned in the introduction critically depended on the essentially *nearest neighbour* nature of the models that were used, which implied that strong locality was exact beyond the second-neighbour distance. The difficulty encountered in using quantum mechanical models in hybrid schemes is due to the non-local effects in the sense of a longer length scale decay in equation (1).

Two groups have published extensively on the use of hybrid schemes in a materials context. Both Abraham *et al* [8] on crack propagation and Ogata *et al* [9] on cracks and surfaces use basically a QM/MM-type approach. (Further examples of QM/MM-type calculations for solids appear in [10].) More recently, Bernstein introduced a modification in the treatment of the quantum zone boundary [11]: using a Green's function technique, he allows a *transition zone* with a thickness of several atomic layers, which is kept in the quantum mechanical calculation, but forces from there are not used in the dynamics. The simulation of a propagating crack in silicon using this technique reproduced the experimental observation that the crack tip is atomically sharp [12]. A similar embedding technique based on Green functions was developed by Inglesfield earlier for static calculations and is reviewed in [13].

3. New scheme

We have recently proposed a novel approach to the handshaking problem, called 'learn on the fly' (LOTF) [5]. While some of the analysis put forward will not be repeated here, we note that the main philosophical difference from the previous approaches is that we do not try to make a combined Hamiltonian from the separate model Hamiltonians; rather we focus on a purely local quantity, the force on each atom. Creating a truly transparent seam between the regions can be broken down into two separate tasks. First, accurate forces must be calculated on every atom in both regions with the appropriate model. Second, the inconsistency of the resultant forces across the boundary has to be resolved, because doing MD simulation directly using forces that come from a variety of different Hamiltonians is inadvisable: for example, the bonds between atoms would not obey an 'action–reaction' principle. The first problem is solved by using a transition zone of finite width (of the order of the characteristic decay length of strong locality). If the quantum mechanical calculation includes atoms not only in the region where we require quantum mechanically accurate forces, but also atoms in the transition zone, the computed forces will be insensitive to the treatment of the boundary of this enlarged region.

The second problem can be addressed in several ways. For example, one could use a weighted average of the forces from the two models in the transition zone, smoothly crossing over from using fully quantum mechanical forces in the quantum region to fully classical forces outside the transition zone. In our scheme, instead of varying the source Hamiltonian for the forces in space, we vary a uniformly valid Hamiltonian in time. We choose a suitably general classical Hamiltonian, and allow its parameters to be different for each atom, bond, angle etc and also allow the parameters to vary in time. So rather than being used directly, the forces that are calculated using the different model Hamiltonians $H_{\text{classical}}$ and H_{quantum} are used to *tune* the parameters of our universal Hamiltonian at each time step: the parameters are varied

to minimize the functional

$$\mathcal{F} = \sum |F_{\text{universal}} - F_{\{\text{classical, quantum}\}}|^2. \quad (2)$$

The MD simulation is then carried out using forces derived from the universal Hamiltonian. Thus the action–reaction principle is restored, and if the tuning is always successful, the resulting forces are extremely close to the separately calculated forces in each region. A very compact way of describing this scheme is that the various models that are assigned to the different parts of the system are being *instantaneously interpolated* by a universal potential. Of course, from a practical point of view, a natural starting point for such a universal potential is an already existing well tested classical potential that gives an adequate description of near-equilibrium configurations for the system at hand.

A few comments are now in order.

- There is an alternative viewpoint from which the above scheme can be understood. Having assumed that we have a classical model that works well in most regions, let us choose this Hamiltonian as the universal one. We want to improve its accuracy in particular places in space and time and we do this by computing relevant new information in those places, using a quantum mechanical model. We then incorporate this new information into our classical trajectories by slightly adjusting the parameters that describe the classical model (only in the region where the relevant new information applies). Whichever viewpoint is more applicable might depend somewhat on one’s personal preferences, but also on just how much the classical parameters had to be readjusted. For small changes, this second viewpoint is very appealing, but if the required alteration of the classical parameters is very drastic and different at every time step, it seems to us more intuitive to talk about interpolation.
- The scheme uses a time-dependent Hamiltonian to carry out the dynamics, and thus energy is not conserved. Although this is not a logical consequence of the above, we have yet to find a formulation which obeys strict energy conservation when the quantum region changes during the simulation, i.e. when atoms regularly cross from the quantum to the classical region and vice versa. In practice a thermostat is used to absorb any small drift in the temperature. This works well except for problems where heat exchange with a bath directly affects what we want to measure, e.g. relative enthalpies.
- The parameter tuning of the universal Hamiltonian need not necessarily be carried out at every time step. One can use the same parameters for a small number of steps, and only then compute the forces with the expensive quantum mechanical model, and then use these forces to retune the parameters. This procedure can be formalized in a predictor–corrector scheme as follows (figure 2 shows a graphical illustration). From a point \mathbf{R}_0 in phase space, having parameters α_0 (let us denote this by (\mathbf{R}_0, α_0)), take a number of MD steps to arrive at point $(\mathbf{R}'_1, \alpha_0)$. Here, quantum calculations and optimization take place, so having calculated new parameters we are at $(\mathbf{R}'_1, \alpha_1)$; this is the end of the ‘predictor’ part. The MD simulation could be continued from this point, but then the trajectory would suffer an abrupt change in its second derivative, because of the discontinuity in the Hamiltonian parameters, α . So instead, we go back to the previous point (\mathbf{R}_0, α_0) and redo the dynamics, interpolating the parameters between α_0 and α_1 . At the end of the interpolation we arrive at a slightly different point in phase space, (\mathbf{R}_1, α_1) , completing the ‘corrector’ part. The interpolation is done using a cubic function with zero derivatives at the two end-points. The cycle could be repeated by recomputing the quantum mechanical forces, getting new parameters and going back to (\mathbf{R}_0, α_0) to reinterpolate. In practice, the points \mathbf{R}'_n and \mathbf{R}_n are very close, and the region of validity of the newly fitted parameters is

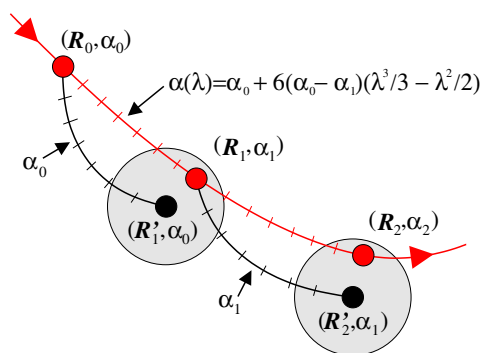


Figure 2. Predictor–corrector-style parameter fitting. The black and red lines represent the predictor and corrector parts of the trajectory, respectively. The grey circles denote the assumed region of validity of the newly fitted parameters at each fitting point (black dots). See the text for a detailed explanation.

large enough that we limit ourselves to one cycle, thereby ensuring a continuous variation of the parameters along the trajectory.

- The alternative viewpoint outlined above highlights a possible computational simplification. Far away from the quantum region, the potential parameters would not be changing very much (or at least the precise details of the trajectories of far-away atoms do not affect processes in the quantum region due to strong locality), so they might as well be left as they are, and not included in the parameter optimization. This speeds up the optimization by a significant factor. If we have several quantum regions which are well separated, the optimizations of the parameters near each region could be carried out in parallel.
- Because of strong locality, the forces on the atoms even in the quantum region can be computed independently. To do this, for each required atom, one carves out a cluster centred on the given atom, and computes the forces, retaining only the force on the central atom (see figure 3(a)). This temporary cluster can be chemically terminated (e.g. with hydrogen atoms) on its surface for improved accuracy (or a smaller cluster radius for any given accuracy). We emphasize that the boundary of these clusters plays a very different role from the boundary of the quantum region which caused most of the difficulties in the traditional embedding schemes. Because each cluster calculation is independent, they can be carried out in parallel requiring minimal bandwidth for inter-processor communication. This possibility makes the scheme particularly suitable for running on massively parallel computer farms, which are typically connected by a comparatively slow communication network. One can always fall back on a single large quantum mechanical calculation, as depicted in figure 3(b). The cluster should then be large enough to include the quantum region and the transition zone, so the forces in the latter can be discarded (i.e. not used in the parameter optimization), thus again avoiding the problems that the introduction of an artificial surface might lead to.

Let us review the approximations that the scheme involves. Firstly, parts of the system are treated with a less accurate classical model. This is a controllable approximation, since the size of the region treated quantum mechanically can be varied at will; this can be done in all embedding schemes. We expect a smooth variation of physical observables as the size of the quantum region is varied because the changing boundary of the quantum region does not bias the calculation: information coming from near the surface of the temporarily carved clusters is

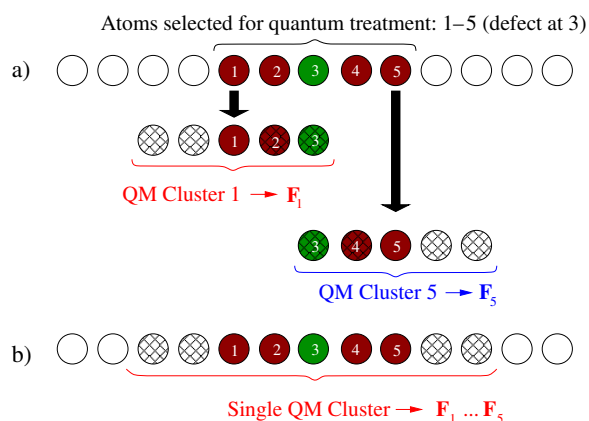


Figure 3. Given a region of interest, e.g. a point defect (green), we declare that this and other atoms within a certain distance should be treated with quantum mechanical accuracy (black). There are two ways of computing the accurate forces that we use in the fitting process, (a) by carving out an independent cluster separately for each QM atom (shown for atoms 1 and 5), and keeping only the force on the central atom from each QM calculation, or (b) by computing a *larger* cluster quantum mechanically, and only keeping the forces on the originally selected QM atoms. In each case, the patterned atoms represent the *transition zone* from which forces are discarded after the QM calculation.

thrown away (see figure 3). Secondly the forces on the QM atoms should be converged with the thickness of the transition zone (or in the case of using independent clusters, with the cluster radius). For example, even in a first-principles *ab initio* quantum mechanical calculation, if we do not expect to get forces more accurately than 0.1 eV \AA^{-1} , it would be a waste of effort to converge to better than this. The same consideration sets the target accuracy of the parameter optimization of our universal Hamiltonian. Below we give typical errors for all these approximations.

We found that about 5–10 MD steps between fitting points in the predictor–corrector dynamics gives a reasonably small deviation from the exact quantum mechanical trajectories. This means that the parameters in the quantum region are essentially continuously retuned, or equivalently, the assumed range of validity of the classical potential with respect to the quantum mechanical model is very small. Outside the quantum region, the trajectories will surely be very different from the quantum mechanical trajectories, but there, by definition, we are only interested in capturing macroscopic behaviour, for which the classical potential is assumed to be adequate.

Although we have lost the ability to do strictly microcanonical simulations, this is offset by other gains. The generality of the force fitting means that any quantum engine can be plugged into the algorithm as a black box, without the need to alter the software. All physical observables can be computed from the MD trajectories, and since averages from different thermodynamical ensembles approach each other in the limit of large samples, long time averages from non-microcanonical MD trajectories are just as good as those from microcanonical runs.

It turns out that the present scheme allows us to make the QM region very small and mobile, i.e. we can continuously select which atoms to treat more accurately during the simulation. This provides the obvious benefit of making the calculations run faster, but highlights the need for robust algorithms for determining where the QM region should be. It is clear that this will depend on the nature of the scientific question one is asking. It is not in general possible

to correct the classical trajectories everywhere they go wrong, since tests show that even at modest displacements from equilibrium, classical models may predict forces that are wrong by as much as $1 \text{ eV } \text{Å}^{-1}$ when compared with quantum mechanical forces.

In practice, this means that, for solid phases, the selection of the quantum region has to be implemented using geometric and topological criteria. For this, we have found it advantageous to keep a running average of atomic positions to filter out the fast optical phonons, and assess the local geometry based on the averaged positions,

$$\tilde{\mathbf{x}}(t) \equiv \frac{1}{1/2 + \tau/\Delta t} \sum_{n=0}^{\infty} e^{-n\Delta t/\tau} \mathbf{x}(t - n\Delta t), \quad (3)$$

where Δt is the time step of the MD simulation and τ is the time constant of the averaging. The averaged coordinates are very well suited to finding the location of point defects and locating bonds that are being broken.

We now present each step of the scheme in order, indicating numerical values for parameters that we have been using. For the sake of concreteness, we take as an example the diffusion of a vacancy in crystalline silicon. This is not a problem which *a priori* stands out as needing an embedded simulation, but precisely because of this, it can provide validation of our approach.

- (i) **Initialization.** Start with a $3 \times 3 \times 3$ cubic unit cell (215 atoms and a vacancy), with atoms near their bulk equilibrium positions, and with a reasonable classical model; in our case we choose the Stillinger–Weber potential [14]. The atomic velocities are randomized and rescaled to the simulation temperature, after which a thermostat is used to maintain an NVT ensemble. We write the classical potential in the following form:

$$E = \sum_i^N \left[\sum_{j \in (i,j)} \epsilon_1 (B_{ij}/r_{ij}^p - A_{ij}) f_{\text{cut}}(r_{ij}) + \sum_{j,k \in (i,j,k)} \epsilon_2 (\cos(\theta_{ijk}) - C_{ijk})^2 f_{\text{cut}}(r_{ij}, r_{ik}) \right]. \quad (4)$$

The cut-off functions f_{cut} go to zero exponentially when the interatomic distance reaches 3.77 Å . Only the parameters A_{ij} , B_{ij} , C_{ijk} are allowed to vary with time; the others and f_{cut} are as in [14].

- (ii) **Extrapolation.** Use the potential with fixed parameters to generate five 1 fs time steps of the system trajectory using the velocity Verlet algorithm.
- (iii) **Testing.** In the latest configuration, the local validity of the classical potential needs to be assessed on a site by site basis, and a selected subset of atoms is flagged for quantum treatment. As discussed above, this selection cannot be made fully automatically. In the present example we are interested in the diffusion of a point defect, so it is natural to describe the region surrounding the defect quantum mechanically. To identify where the vacancy is, we use equation (3) with a time constant of 100 fs, and match the averaged atomic coordinates with their nearest lattice points. Having located the site of the missing atom, we flag all atoms that are within 7.5 Å of its lattice point (about 100 atoms).
- (iv) **Quantum mechanical calculations.** We compute accurate forces on all flagged atoms using an empirical tight binding method [6] and the independent cluster approach described above. The cluster radius is set to 7 Å ; this yields about 150 atoms in each cluster, including terminating hydrogens. The termination atoms are put in place of the Si atoms that lie just outside the given cluster and their positions are adjusted to the equilibrium Si–H distance along the direction of each Si–Si bond that was cut. The Hellmann–Feynman force on the central atom is then computed after a direct diagonalization of the Hamiltonian. This force is shown to converge quickly with cluster radius (see figure 1) and to be relatively

insensitive to the precise termination strategy. We determined in a separate test (0.5 ps LOTF simulation of a vacancy diffusion event at 1400 K) that a 7 Å radius is large enough that the RMS deviation from the exact TB forces (computed with periodic boundary conditions) is $<0.1 \text{ eV \AA}^{-1}$. Constructing clusters of similar size including all Si atoms up to a chosen neighbour shell rather than up to a fixed radius yields analogous results.

- (v) **Force fitting.** We tune the parameters of the classical potential until it reproduces the accurate forces, by minimizing (2) using the conjugate gradients technique. The starting values for the parameters are always reset to be the same as those of the original unmodified potential. For atoms that do not have a quantum mechanical force calculated on them, the current classical force is used as the target of the optimization. Note that we require the evaluation of the classical potential for a large number of different parameter sets, all at the same atomic configuration. We have devised special routines that reformulate the potential to be of the form

$$E = W^0(\mathbf{R}) \cdot \mathbf{1} + W^1(\mathbf{R}) \cdot \alpha + W^2(\mathbf{R}) \cdot \alpha^2,$$

where only the vectors W^n depend on the atomic coordinates and are computed once before the optimization starts; α and α^2 indicate the vector of the parameters and their squares, respectively. The forces can be written in a similar way and thus during the optimization, most of the CPU time is spent in BLAS routines. Tests show that parameters related to atoms which are further than a few neighbour shells outside the QM region do not change appreciably with the fit. The optimization is therefore limited to the terms of the potential involving atoms in a spherical region of 12 Å radius centred on the diffusing defect. About 30 conjugate gradient steps are sufficient for performing a converged fit for the silicon vacancy system, the maximum residual deviation from the target forces being typically less than 0.01 eV \AA^{-1} .

- (vi) **Interpolation.** Now return the state of the system to that before the extrapolation and rerun the dynamics, interpolating the potential parameters between the old and the new values. We use a cubic spline interpolation, so the time derivatives of the parameters are zero at the fitting points.
- (vii) **Return to 2.**

4. Validation

Using this procedure, we computed the diffusivity for the Si vacancy [15] by running a 1 ns constant temperature (CT) simulation of a 215-atom periodic system at $T = 1400 \text{ K}$ and obtained a value of $(2.1 \pm 0.4) \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$. This compares well with the value of $(2.3 \pm 0.2) \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$, calculated applying the (same) tight binding scheme [6] to the full system of 215 atoms and the vacancy. For comparison, the Stillinger–Weber potential gives $(3.4 \pm 1.0) \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$, a much larger value. It must be stressed that the diffusivity is a quantitative test of the embedding method, and should not be taken to be an application where embedding is really necessary. A 215-atom silicon system is most efficiently simulated by standard techniques.

This is even more true of our next test, where we look at the temperature dependence of the vacancy diffusivity. To enable the large number of runs that a good comparison requires, we have chosen to use a $2 \times 2 \times 2$ cubic unit cell with 63 atoms. Here, the quantum forces used for the parameter fitting in the LOTF run were calculated using the fully periodic system, rather than with clusters.

Figure 4 shows our results for the vacancy diffusivity in Si as a function of temperature, obtained by CT simulations using the fixed-parameter SW potential, two different TB

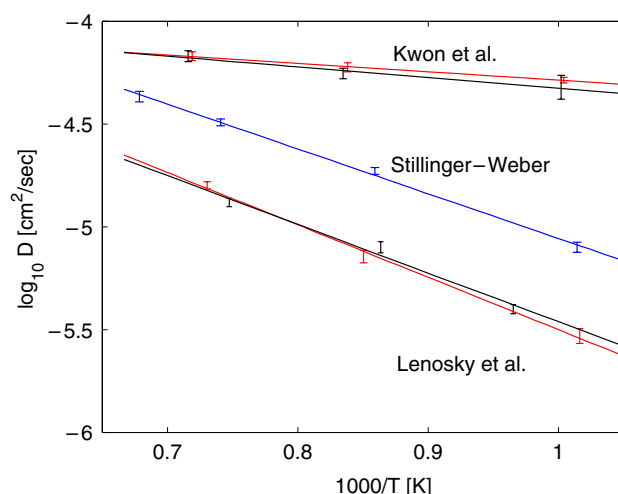


Figure 4. Arrhenius plot of the silicon vacancy diffusivity obtained by the present scheme fitted ‘on the fly’ on different TB models (red), compared with the results of fully TB simulations (black) and fixed-parameter SW potential (blue).

schemes [16, 17] and our hybrid LOTF scheme. We note that the large difference between the results obtained by fitting the scheme on the two different quantum models is due to the different predictions of these models in accurate quantitative computations, which are faithfully reproduced by the hybrid method. This emphasizes the fact that the present scheme can, at best, be expected to reproduce the results of the QM model that it is given, but can in no way improve its accuracy.

A different quantitative test is provided by the example of stressing a silicon nanobar, shown in figure 5. Starting with a $2 \times 2 \times 10$ supercell of the cubic unit cell, we introduce a vacancy and relax the structure to equilibrium. Then the atomic coordinates are rescaled parallel to the bar axis, and again relaxed to equilibrium while holding the ends fixed. In the hybrid simulation, the QM region included atoms within 4 \AA of the vacancy measured along the bar. In this case, rather than using clusters, the tight binding forces were calculated by considering a cluster that was a bit larger than the QM region (by 5 \AA on each side), and again terminated with hydrogen atoms where bonds were broken in the construction of the cluster.

Figure 5(b) shows the resulting stress–strain curves (the sum of the z components of the forces on the fixed atoms at one end of the bar is proportional to the stress). The hybrid simulation and the fully quantum simulation give the same value for the critical stress (dashed line) at which the bar fails mechanically near the vacancy, while the critical stress computed classically is significantly higher. This example also illustrates that the microscopic trajectories are, of course, not necessarily reproduced: the curves have minor jogs and kinks due to small changes in the surface reconstruction of the nanobar. However, these only affect the critical strain and not the value of the critical stress, which is controlled by the major mechanical failure near the vacancy; this is what we set out to model quantum mechanically.

5. Brittle failure

Having validated our scheme in small systems where it was possible to compute observables using a fully quantum mechanical method, we now turn to a system which has been epitomized

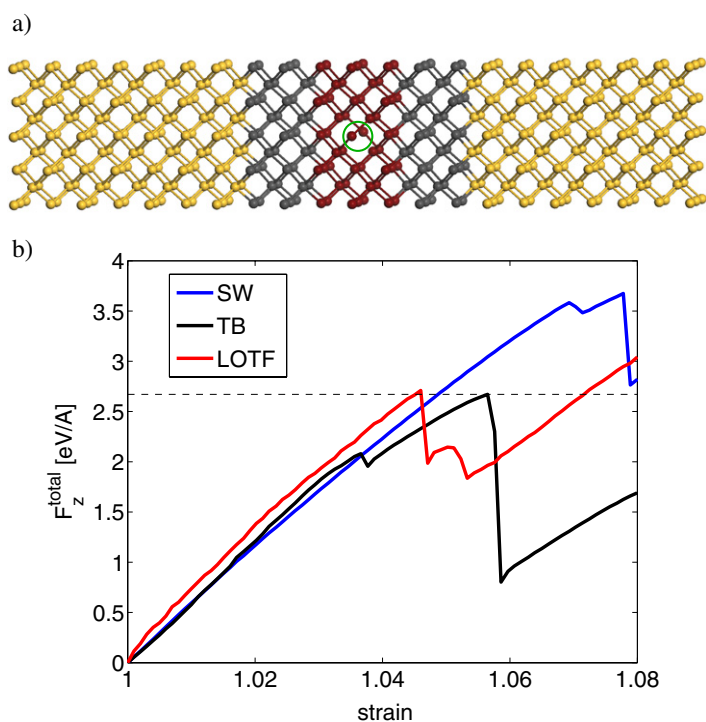


Figure 5. (a) Geometry of a silicon nanobar with a vacancy (green circle) before relaxation. The forces on the red atoms are obtained from a TB calculation on a single cluster that also includes the grey atoms. (b) Stress–strain curves of the nanobar, calculated using three different methods: Stillinger–Weber (blue), tight binding (black) and LOTF (red).

as the benchmark for hybrid schemes, brittle failure of covalent crystals. When single crystals of silicon or diamond are strained beyond breaking point, they relieve the stress by opening cracks. A universal observation is that, below the brittle to ductile transition temperature, these cracks are atomically sharp and the resulting newly opened surfaces are almost atomically flat [18]. The system is evidently multiscale, and the behaviours on the different length scales are highly correlated: the large stress concentration near the crack tip induces the breakage of atomic bonds, and it is the successively broken bonds that allow the crack tip (and the stress field) to advance forward as the crack propagates.

To date, all purely classical simulations have failed to reproduce a basic feature of the system, namely that the crack tips remain atomically sharp as the crack opens. This is principally due to the systematic overestimation of lattice trapping effects [12], which results in an excess of applied stress with respect to the critical value for getting the crack moving in the first place. This excess of elastic energy is dissipated by tip blunting and local melting.

Other efforts at developing hybrid schemes have also concentrated on cracks [8, 9]. The principal advantage of our scheme is that the quantum region can be made very small (300–600 atoms for a silicon crack tip) and is allowed to follow the tip as the crack propagates, resulting in a significant computational speed-up without compromising the accuracy of the calculation.

Figure 6 shows four snapshots of our simulation of the $(111)[\bar{1}\bar{1}0]$ crack system, as it propagates in the $[11\bar{2}]$ direction at 0 K. We used the *ab initio* SIESTA [19] package to treat

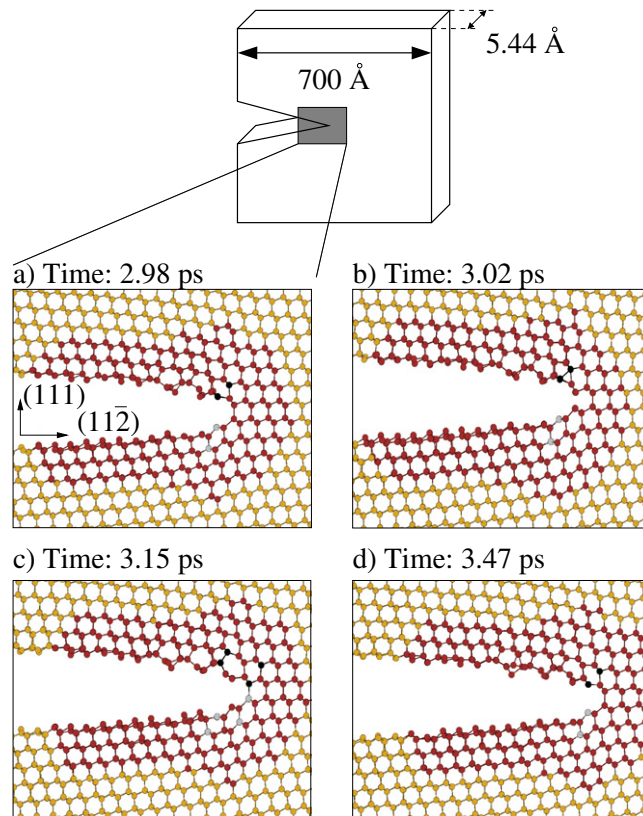


Figure 6. Four snapshots of the opening $(111)[\bar{1}\bar{1}0]$ crack system simulated using LOTF. Atoms in red were treated quantum mechanically using the *ab initio* SIESTA package. The quantum region follows the crack tip as it moves from left to right. Note the pentagons and heptagons of the 2×1 Pandey reconstruction on the top surface. See the text for an explanation of the dynamical aspects.

atoms in the moving quantum region. In addition to showing smooth atomically flat surfaces opening, using such a sophisticated QM engine revealed that the surface reconstructions are different on the upper and lower surfaces, with the upper surface showing the 2×1 Pandey reconstruction, as indicated by the alternating pentagons and heptagons. The reason that only one surface shows this reconstruction can be understood by considering the dynamics of atoms near the crack tip. The atoms, shown in black, that were part of a hexagon are forced towards each other as the crack tip passes (see panels (a) and (b)), thus driving the formation of a pentagon. In panel (d), the crack has moved and the next pentagon is about to form. The corresponding atoms on the lower surface (shown in grey) are forced away from each other, preventing pentagon formation. Further results on the crack propagation problem in silicon will be published elsewhere.

6. Conclusion

At present, hybrid simulation techniques are not yet used very widely. This stems not only from the fact that the surge of interest in method development is very recent, but also from the technical difficulties in properly representing the boundary between the regions described

by different models. A crucial element needed for the future success of hybrid schemes is generality and a ‘black box’ nature: the linkage technique should not be tied to any particular quantum mechanical or classical atomistic method, but the users should be able to use any model they wish, with relative ease, without having to rewrite, retest and revalidate the combined scheme. In our view, having an extended transition zone between the different regions is essential for a seamless coupling with *controllable* accuracy. The decision of whether to bother with the complexity of a hybrid methodology in treating any specific problem depends on the fraction of atoms that will require accurate treatment. While the precise crossover value will depend very much on the details (e.g. cost and speed ratio of the various different models, the overall required accuracy), we can say with confidence that problems where the effective dimensionality of an active process that is critical for the simulation is lower than that of the whole region which is needed to describe the system are good candidates. Most defect-driven processes fall in this category.

Acknowledgments

The authors acknowledge funding from the EPSRC Grant GR/S61263/01 and EU Grant HPRN-CT-2000-0154. We have had useful and stimulating discussions with Noam Bernstein, Daan Frenkel, Stefan Goedecker and David Pettifor.

References

- [1] Nieminen R M 2002 *J. Phys.: Condens. Matter* **14** 2859
- [2] Martonak R *et al* 2000 *Phys. Rev. Lett.* **84** 682
- [3] Kohlhoff S *et al* 1991 *Phil. Mag. A* **64** 851
- [4] Tadmor E B *et al* 1996 *Phil. Mag.* **73** 1529
Tadmor E B *et al* 1996 *Langmuir* **12** 4529
Shenoy V B *et al* 1998 *Phys. Rev. Lett.* **80** 742
Knap J and Ortiz M 2001 *J. Mech. Phys. Solids* **49** 1899
- [5] Csányi G *et al* 2004 *Phys. Rev. Lett.* **93** 175503
De Vita A and Car R 1998 *Mater. Res. Soc. Symp. Proc.* **491** 473
- [6] Bowler D R *et al* 1998 *J. Phys.: Condens. Matter* **10** 3719
- [7] 2003 Combined QM/MM calculations in chemistry and biochemistry *THEOCHEM* **632** (special issue)
- [8] Abraham F F *et al* 1998 *Comput. Phys.* **12** 538
Broughton J Q *et al* 1999 *Phys. Rev. B* **60** 2391
Abraham F F *et al* 2000 *Mater. Res. Soc. Bull.* **25** 27
- [9] Ogata S *et al* 2001 *Comput. Phys. Commun.* **138** 143
Ogata S *et al* 2002 *Comput. Phys. Commun.* **149** 30
Ogata S *et al* 2004 *J. Appl. Phys.* **95** 5316
Ogata S and Belkada R 2004 *Comput. Mater. Sci.* **30** 189
- [10] Du M-H and Cheng H-P 2003 *Int. J. Quantum Chem.* **93** 1
Du M-H *et al* 2003 *J. Chem. Phys.* **119** 6418
Ellis D E and Warschkow O 2003 *Coord. Chem. Rev.* **238** 31
- [11] Bernstein N 2001 *Europhys. Lett.* **55** 52
- [12] Bernstein N and Hess D W 2003 *Phys. Rev. Lett.* **91** 025501
- [13] Inglesfield J E 2001 *Comput. Phys. Commun.* **137** 89
- [14] Stillinger F and Weber T 1985 *Phys. Rev. B* **31** 5262
- [15] Tang M *et al* 1997 *Phys. Rev. B* **55** 14279
- [16] Kwon I *et al* 1994 *Phys. Rev. B* **49** 7242
- [17] Lenosky T J *et al* 1997 *Phys. Rev. B* **55** 1528
- [18] Lawn B 1993 *Fracture of Brittle Solids* (Cambridge: Cambridge University Press)
- [19] Sanchez-Portal D *et al* 1997 *Int. J. Quantum Chem.* **65** 453